THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2005/33

Sieve Nonparametric Likelihood Methods for Unit Root Tests

by

Francesco Bravo

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Sieve nonparametric likelihood methods for unit root tests*

Francesco Bravo†
University of York

Abstract

This paper develops a new test for a unit root in autoregressive models with serially correlated errors. The test is based on the "empirical" Cressie-Read statistic and uses a sieve approximation to eliminate the bias in the asymptotic distribution of the test due to presence of serial correlation. The paper derives the asymptotic distributions of the sieve empirical Cressie-Read statistic under the null hypothesis of a unit root and under a local-to-unity alternative hypothesis. The paper uses a Monte Carlo study to assess the finite sample properties of two well-known members of the proposed test statistic: the empirical likelihood ratio and the Kullback-Liebler distance statistic. The results of the simulations seem to suggest that these two statistics have, in general, similar size and in most cases better power properties than those of standard Augmented Dickey-Fuller tests of a unit root.

The paper also analyses the finite sample properties of a sieve bootstrap version of the (square of) the standard Augmented Dickey-Fuller test for a

*Financial support from ESRC under research grant no. RES-000-22-0980 is gratefully acknowledged.

†Aspects of this research were presented at the CeMMAP workshop on empirical likelihood methods at IFS-UCL February 2005, and the Econometric Study Group annual meeting at the University of Bristol July 2005. I would like to thank Whitney Newey, Robert Taylor and Andy Tremayne for some useful comments and suggestions. All remaining errors are my own responsibility.
Address correspondence to: Francesco Bravo, Department of Economics and Related Studies, University of York, York, YO10 5DD, United Kingdom. email fb6@york.ac.uk

unit root. The results of the simulations seem to indicate that the bootstrap does solve almost completely the size distortion problem, yet at the same time produces a test statistic that has considerably less power than either that of the empirical likelihood or of the Kullback-Liebler distance statistic.

Keywords and Phrases. Autoregressive approximation, Bootstrap, Empirical Cressie-Read statistic, Generalized Empirical likelihood, Linear process, Unit root test.

1 Introduction

Asymptotic theory for testing for the presence of a unit root in autoregressive models, the so-called unit root hypothesis, is well developed - see Phillips and Xiao (1998) for an updated survey- and routinely applied in macroeconomics and finance where most time series display unit root characteristics. There exists however a large body of Monte Carlo evidence showing that the conventional asymptotic approximations to the null sampling distribution of most test statistics for a unit root are inaccurate, particularly in the case of autoregressive models with weakly dependent errors. As a result the exact levels of these tests are often very different from nominal levels derived from asymptotic theory. In such cases one option is to use the bootstrap to improve the quality of the asymptotic approximation. The validity of bootstrap methods for unit root models was originally investigated by Basawa, Mallikand, McCormick, Reeves and Taylor (1991) and more recently by Psaradakis (2001), Romano and Wolf (2001), Park (2002), and Chang and Park (2003) among others. Another option is to consider alternative tests for a unit root. In this paper we follow the latter and propose a new test for a unit root based on the so-called empirical[1] Cressie-Read method.

The empirical Cressie-Read method, introduced by Baggerly (1998) as a generalization of Owen's (1988) empirical likelihood, provides a nonparametric likelihood-based alternative to bootstrap procedures for inference in a number of nonparametric situations. The ECR discrepancy produces a very large family of nonparametric likelihood ratio-type statistics which includes the empirical likelihood ratio (Owen, 1988), the Kullback-Liebler distance (DiCiccio and Romano, 1990) based on Efron (1981) nonparametric tilting method, the Euclidean likelihood ratio (Owen, 1991), and the least favorable family-based likelihood ratio (Lee and Young, 1999) among others. As noted by Baggerly (1998), all members of the ECR family enjoy a number of desirable statistical properties: they yield convex confidence regions (at least for a multivariate mean) whose shape is typically data-determined. Furthermore ECR regions are transformation invariant, do not require estimation of scale and, as recently shown by Bravo (2005*a*), they all admit Bartlett-type corrections that can lead to improved

[1]We use the term "empirical" to emphasise the central role played by the empirical distribution of the data

inference. The ECR method can also be used to construct new estimators and test statistics that are asymptotically equivalent to those based on standard asymptotic methods, but are characterized by better finite sample properties (either in terms of bias, and/or in terms of accuracy and power). For example Imbens and Spady (2002) showed that all members of the ECR family give rise to alternative estimators and associated statistics to those based on Hansen's (1982) generalized method of moments (GMM) method but with better finite sample properties. For all these reasons, it is perhaps not surprising that the nonparametric likelihood approach to inference based on the ECR method has attracted recently a great deal of interest in econometrics and statistics.

In this paper we show how the ECR method can be used to obtain a nonparametric likelihood based family of tests for a unit root in an autoregressive model with errors parameterized as a general linear process. The results of the paper are based on the same autoregressive approximation used by Chang and Park (2002) in the context of Augmented Dickey-Fuller (ADF) unit root tests, and by Psaradakis (2001) and Chang and Park (2003) in the context of bootstrap unit root tests. This approximation captures the dependent structure of the errors by fitting autoregressive models with order increasing with the sample size at an appropriate rate. Since the approximation is sometimes known in time series literature as sieve approximation (see e.g. Bühlmann (1997)), we shall call the resulting ECR method sieve ECR (SECR henceforth).

In this paper we establish the asymptotic distributions of the SECR statistic under the null hypothesis of a unit root and under a sequence of local-to-unity alternatives. The resulting distributions correspond to the square of well-known functionals of a Brownian motion and Ornstein-Uhlenbeck process, respectively. These results are of theoretical interest and generalize (and/or) complement results obtained by Chuang and Chan (2002), Bravo (2005*b*) and Bravo (2005*c*). We note here that to derive these results we rely on methods developed and used by Donald, Imbens and Newey (2003) and by Bravo (2005*c*).

In this paper we also use simulations to assess the finite sample properties of the two most well-known (and used in practice) members of the SECR statistic, namely the empirical likelihood ratio and the Kullback-Liebler distance. Their finite sample properties are compared with those of a standard (squared) ADF t- statistic, and with those based on the sieve bootstrap (squared) ADF t- statistic. The results of the simulations are encouraging and seem to suggest that in a number of cases of practical relevance the SERC method can produce test statistics with finite sample properties that compare favorably with those of the original and bootstrapped (squared) ADF t- statistic.

The rest of the paper is structured as follows: next section reviews the ECR method for inference in the case of independent and identically distributed observations. Section 3 introduces the sieve approximation, and shows how it can be used in the context of testing for a unit root in a simple autoregressive model of order one. Section 4 contains the main results of the paper. Section 5 reviews the sieve bootstrap method in the context of unit roots tests. Section 6 contains the results of a Monte Carlo study, while Section 7 contains some concluding remarks and indications for future research. An Appendix contains all the proofs.

2 ECR method

Like empirical likelihood, the empirical Cressie-Read (henceforth ECR) method uses numerical optimization to estimate the unknown distribution of the data subject to a given restriction assumed to contain all the information available in the sample. The resulting constrained estimator can be used to make inference about the restriction itself, using the well-known fact that without restriction the empirical distribution function is an optimal estimator (i.e. it is the maximum nonparametric likelihood estimator) of the unknown distribution of the data. Specifically, the *discrepancy* (or distance) between the constrained and unconstrained (that is the empirical distribution function) estimators of the distribution of the data can be used to assess whether the imposed restriction is supported by the data. In the case of ECR method the discrepancy is measured by the Cressie-Read power-divergence statistic (Read and Cressie, 1988). For alternative, more general, discrepancies see Smith (1997) and Corcoran (1998).

Suppose that the observations $(x_i)_{i=1}^n$ are independent identically distributed (i.i.d.) $\Re^q$-valued random vectors from an unknown distribution F, let $\theta \in \Theta \subseteq \Re^q$ be an unknown parameter vector associated with F, and E denote the expectation operator with respect to F. We assume that the information about F and θ is available in the form

$$E\left[g\left(z,\theta_0\right)\right]=0, \tag{1}$$

for some specified value θ_0 of θ, with $g\left(z,\theta\right):\Re^q\times\Theta\rightarrow\Re^q$-valued vector of known functionally independent functions. The typical approach to inference for θ in (1) is to use the analogy principle and define the Z-estimator $\widehat{\theta}$ solution to the empirical analogue of (1), *viz.*

$$g_n\left(\widehat{\theta}\right):=\sum g\left(z_i,\widehat{\theta}\right)/n=0. \tag{2}$$

Then tests and confidence regions for θ can be based on the score-type statistic

$$ng_n\left(\widehat{\theta}\right)'\left[\sum g\left(z_i,\widehat{\theta}\right)g\left(z_i,\widehat{\theta}\right)'/n\right]^{-1}g_n\left(\widehat{\theta}\right)'$$

since the latter converges to a χ^2_q under mild regularity conditions. Alternatively one can use the ECR to obtain nonparametric likelihood-based inferences for θ. Specifically let w_i denote the weight that F places on the ith observation x_i, where F is an arbitrary probability measure on $\Re^p$, and let $\widehat{w}_i=1/n$ denote the weight (probability mass) that the empirical distribution function F_n assigns to each of the observations. The ECR discrepancy between w_i and $1/n$ (that is between F and F_n) is given by

$$CR\left(w_i,1/n,\gamma\right)=\sum_{i=1}^{n}\left[\left(nw_i\right)^{-\gamma}-1\right]/\left[\gamma\left(\gamma+1\right)\right]$$

where $\gamma\in\Re$ is a user-specified parameter. In particular for $\gamma=-2$ one obtains the Euclidean likelihood, for $\gamma=-1$ the Kullback-Liebler, and for $\gamma=0$ the empirical likelihood ratio statistics[2]. The ECR method chooses the unknown weights w_i's so that the null hypothesis $H_0:\theta=\theta_0$ holds in the sample and the Cressie-Read statistic is minimized. To be specific the w_i's solve the following constrained minimizations

$$\min_{w_i}\left\{CR\left(w_i,1/n,\gamma\right)\mid\sum_{i=1}^{n}w_i=1,\sum_{i=1}^{n}w_i g\left(x_i,\theta_0\right)=0\right\}. \tag{3}$$

Assume that

$$\Pr\left\{0\in ch\left(g\left(x_1,\theta_0\right),...,g\left(x_n,\theta_0\right)\right)\right\}=1 \text{ as } n\rightarrow\infty,$$

[2]Note that the two degenerate cases $\gamma=-1$ and $\gamma=0$ are handled by taking the limits.

that is, asymptotically the true mean of the distribution lies within the convex hull $ch(\cdot)$ of $g(x_1, \theta_0), ..., g(x_n, \theta_0)$, for otherwise (3) does not admit a solution since it is impossible to reweight the data so that the mean is 0. A Lagrange multiplier argument shows that the unique solution to (3), is given by

$$\begin{aligned}\widetilde{w}_i(\theta_0, \gamma') &= n^{-1}\left(1 + \widehat{\zeta} + \widehat{\xi}' g(x_i, \theta_0)\right)^{-1/(1+\gamma')} \quad \text{for } \gamma' \in \Re/\{-1, 0\}, \\ \widetilde{w}_i(\theta_0, -1) &= \widehat{\zeta}\exp\left(\widehat{\xi}' g(x_i, \theta_0)\right), \quad \widetilde{w}_i(\theta_0, 0) = n^{-1}\left(1 + \widehat{\xi}' g(x_i, \theta_0)\right)^{-1} \quad (4)\end{aligned}$$

where the Lagrange multipliers $\widehat{\zeta} \in \Re$ and $\widehat{\xi} \in \Re^q$ are determined by the constraints $\sum_{i=1}^{n} w_t = 1$ and $\sum_{i=1}^{n} w_i g(x_i, \theta_0) = 0$, respectively. Using (4) in (3) it follows that the ECR statistic for θ_0 is given by $W(\theta_0, \gamma) = 2CR(\widetilde{w}_t(\theta_0, \gamma), 1/n, \gamma)$ where

$$\begin{aligned}W(\theta_0, \gamma') &= \frac{2}{\gamma'(\gamma'+1)}\sum_{t=1}^{n}\left[\left(1 + \widehat{\zeta} + \widehat{\xi}' g(x_i, \theta_0)\right)^{\gamma'/(\gamma'+1)} - 1\right], \quad \text{for } \gamma' \in \Re/\{-1, 0\} \\ W(\theta_0, -1) &= 2\sum_{t=1}^{n}\left[n\widehat{\zeta}\exp\left(\widehat{\xi}' g(x_i, \theta_0)\right)\right]\log\left[n\widehat{\zeta}\exp\left(\widehat{\xi}' g(x_i, \theta_0)\right)\right], \\ W(\theta_0, 0) &= 2\sum_{t=1}^{n}\log\left(1 + \widehat{\xi}' g(x_i, \theta_0)\right), \quad (5)\end{aligned}$$

and $W(\beta_0, -1)$, $W(\beta_0, 0)$ correspond to the Kullback-Liebler and the empirical likelihood ratio statistics, respectively. A straightforward modification to Baggerly's (1998) arguments shows that $W(\theta_0, \gamma) \xrightarrow{d} \chi_q^2$, which in turn may be used to build nonparametric likelihood based confidence regions for θ by inversion.

3 Sieve approximation for unit root tests

Consider the following $AR(1)$ process

$$\begin{aligned}y_t &= \beta y_{t-1} + u_t, \quad t = 1, 2, ..., n \quad (6) \\ u_t &= \Psi(L)\varepsilon_t = \sum_{j=1}^{\infty}\psi_j\varepsilon_{t-j}\end{aligned}$$

where, for simplicity, $y_0 = 0$[3], and the innovations ε_t form a martingale difference sequence satisfying the following assumptions:

[3] The initial value of y_t does not affect the asymptotic results obtained below as long as $y_t = O_p(1)$.

M (I) $E\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right) = \sigma^2$ $a.s.$, (II) $E\left(|\varepsilon_t|^\delta|\mathcal{F}_{t-1}\right) < \infty$ $a.s.$ for some $\delta \geq 4$,

and $\Psi(L)$ satisfies

LP $\Psi(z) \neq 0$ for all $|z| \leq 1$ and $\sum_{j=1}^{\infty} j^k |\psi_j| < \infty$ for some $k \geq 1$.

Assumptions **M** and **LP** imply that u_t is a strictly stationary linear process driven by martingale difference innovations. Note also that **LP** includes models with polynomial decay of the coefficients $\{\psi_j\}_{j=0}^{\infty}$. General $ARMA(p,q)$ processes satisfy **LP** with an exponential decay of $\{\psi_j\}_{j=0}^{\infty}$.

It is well known (see, for example, Phillips (1987a)) that the dependent structure of the error process u_t introduces a bias term that affects the asymptotic distributions of the t- statistic or coefficient test for the unit root hypothesis $H_0 : \beta = 1$ in (6). One way to solve this problem is to correct the test statistics with nonparametric estimates of the bias as originally proposed by Phillips (1987a) and Phillips and Pierron (1988)[4]. Alternatively one can approximate u_t by a finite autoregressive process of order increasing with the sample size as recently suggested by Chang and Park (2002). As mentioned in the Introduction, this autoregressive approximation is sometimes known in time series literature as sieve approximation (Bühlmann, 1997), since it is based on a sequence of autoregressive processes (i.e. a sequence of finite dimensional parametric models) approximating a linear process, which can be thought of as an infinite dimensional nonparametric model. It should be noted that in the context of tests for a unit root, a similar idea was originally proposed by Said and Dickey (1984) and more recently by Xiao and Phillps (1998) as an extension to the traditional augmented Dickey-Fuller (ADF) test for a unit root with weakly dependent errors. Both Said and Dickey (1984) and Xiao and Phillps (1998) however considered only linear processes with geometrically decaying coefficients and i.i.d. innovations, as opposed to the general linear processes considered in Chang and Park (2002) and in this paper.

Let

$$\Phi(L)\, u_t = \varepsilon_t \tag{7}$$

[4]Note that the corrections proposed by Phillips (1987a) and Phillips and Pierron (1988) are also valid for other weakly dependent structures of the errors. In particular they are valid if u_t is assumed to be a strong-mixing process with mixing coefficients α_m satisfying $\sum_{j=1}^{\infty} \alpha_j^{1-2/\alpha} < \infty$ for some $\alpha > 2$.

denote the $AR(\infty)$ representation of the linear process u_t, and notice that **LP** implies that $\Phi(z)$ is bounded away from 0 for $|z| \leq 1$ and that $\sum_{j=1}^{\infty} j^k |\phi_j| < \infty$. Consider the following $AR(p)$ approximation to u_t

$$u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_{p,t} \tag{8}$$

where $\varepsilon_{p,t} = \varepsilon_t + \sum_{j=p+1}^{\infty} \phi_j u_{t-j}$, and note that by iterated expectations, **M(II)** and **LP** $E|\varepsilon_{p,t} - \varepsilon_t|^\gamma = o(p^{-\gamma k})$ (Chang and Park, 2002), implying that the larger p becomes, the smaller is the error in the autoregressive approximation (8). The basic idea in the sieve approximation for u_t is to allow the order p of the autoregressive approximation to increase at an appropriate rate with the sample size, that is $p = p(n) \to \infty$ as $n \to \infty$. Using (8), we can rewrite (6) as

$$y_t = \beta y_{t-1} + \sum_{j=1}^{p(n)} \phi_j \Delta y_{t-j} + \varepsilon_{p(n),t}, \tag{9}$$

where $\Delta y_t = u_t$. Define the sieve (or augmented) "score" function

$$m_{p(n),t}(\beta, \phi) = \begin{bmatrix} y_{t-1} & \Delta y_{t-1} & \dots & \Delta y_{t-p(n)} \end{bmatrix}' \left(y_t - \beta y_{t-1} - \sum_{j=1}^{p(n)} \phi_j \Delta y_{t-j} \right), \tag{10}$$

and let $\begin{bmatrix} \widehat{\beta} & \widehat{\phi}' \end{bmatrix}'$ denote the least square estimator that solves $\sum_{t=1}^{n} m_{p(n),t}\left(\widehat{\beta}, \widehat{\phi}\right) = 0$. Then using (9) the unit root hypothesis $H_0 : \beta = 1$ can be tested using the ADF t-statistic

$$ADF_t = \left(\widehat{\beta} - 1\right) / \widehat{\sigma}_{\widehat{\beta}} \tag{11}$$

where $\widehat{\sigma}_{\widehat{\beta}}$ is the standard error for the estimated coefficient $\widehat{\beta}$.

Let $B(r)$ denote a standard Brownian motion on $C[0,1]$, the space of continuous functions on the interval $[0,1]$, and $\Rightarrow$ denote weak convergence in distribution. Chang and Park (2002) show that under the minimal rate condition[5] $p(n) = o(n^{1/2})$ the asymptotic distribution of (11) is

$$ADF_t \Rightarrow \left(\int_0^1 B(r)\, dB(r) \right) / \left(\int_0^1 B^2(r)\, dr \right)^{1/2},$$

that is the distribution of ADF_t coincides with that of the t-statistic for a unit root obtained by Dickey and Fuller (1979) in the absence of serially correlated errors.

[5] Note however that this rate is not sufficient for consistency of the estimates of the ϕ_j's. See Chang and Park (2002) for more details.

4 Main results

If the error process u_t in (6) was a martingale difference sequence, the martingale property of the score function $m_t(\beta) = y_{t-1}(y_t - \beta y_{t-1})$ under the true value of β, say β_0, implicitly imposes the moment restriction $E[m_t(\beta_0)] = 0$ that, as shown by Bravo (2005*b*), can be used to build an ECR statistic for the hypothesis $H_0 : \beta = \beta_0$. In the case of weakly dependent errors however the same restriction does not hold, and thus cannot be used by the ECR method. On the other hand, under $H_0 : \begin{bmatrix} \beta & \phi' \end{bmatrix}' = \begin{bmatrix} \beta_0 & \phi_0' \end{bmatrix}'$

$$E\left[E\left(m_{p(n),t}(\beta_0, \phi_0) \mid \mathcal{F}_{t-1}\right)\right] \to 0 \quad \text{as } p(n) \to \infty,$$

that is the sieve score $m_{p(n),t}(\beta_0, \phi_0)$ gives rise to a set of $p(n)$ approximate moment restrictions that can be tested using ECR method as described in Section 2. In practice the resulting sieve ECR (SERC) finds the w_i's consistent with $E\left[m_{p(n),t}(\beta_0, \phi_0)\right] = 0$ by solving

$$\min_{w_t} \left\{ CR(w_t, 1/n, \gamma) \mid \sum_{t=1}^{n} w_t = 1, \sum_{t=1}^{n} w_t m_{p(n),t}(\beta_0, \phi_0) = 0 \right\}. \tag{12}$$

As discussed in Section 2, as long as

$$\Pr\left(0 \in ch\left\{ \begin{matrix} m_{p(n),1}(\beta_0, \phi_0) & \cdots & m_{p(n),t}(\beta_0, \phi_0) \end{matrix} \right\}\right) \to 1 \text{ as } n \to \infty,$$

a unique solution to (12) exists and is given as in (4). Then the SERC statistic for the null hypothesis $H_0 : \begin{bmatrix} \beta & \phi' \end{bmatrix}' = \begin{bmatrix} \beta_0 & \phi_0' \end{bmatrix}'$ is given as in (5), that is

$$\begin{aligned} W(\beta_0, \phi_0, \gamma') &= \frac{2}{\gamma'(\gamma'+1)} \sum_{t=1}^{n} \left[\left(1 + \widehat{\zeta} + \widehat{\xi}' m_{p(n),t}(\beta_0, \phi_0)\right)^{\gamma'/(\gamma'+1)} - 1 \right] \\ \text{for } \gamma' &\in \Re / \{-1, 0\}, \\ W(\beta_0, \phi_0, -1) &= 2 \sum_{t=1}^{n} \left[n\widehat{\zeta} \exp\left(\widehat{\xi}' m_{p(n),t}(\beta_0, \phi_0)\right) \right] \log \left[n\widehat{\zeta} \exp\left(\widehat{\xi}' m_{p(n),t}(\beta_0, \phi_0)\right) \right], \\ W(\beta_0, \phi_0, 0) &= 2 \sum_{t=1}^{n} \log\left(1 + \widehat{\xi}' m_{p(n),t}(\beta_0, \phi_0)\right). \end{aligned} \tag{13}$$

The unit root hypothesis $H_0 : \beta = 1$ is a composite hypothesis with β the parameter of interest and ϕ a vector of nuisance parameters. Like empirical likelihood, ECR deals with nuisance parameters by profiling (i.e. concentrating them out). Let

$$W\left(1,\widehat{\phi},\gamma\right) := 2\min_{\phi} W\left(1,\phi,\gamma\right)$$

denote the profile SECR for $H_0 : \beta = 1$. In addition to **M**(I),(II) and **LP** we assume that

MU $\sup_t E\left(|\varepsilon_t|^{\alpha} |\mathcal{F}_{t-1}\right) < \infty$ $a.s.$ for some $\alpha > 2$,

and that as $p(n) \to \infty$

P (I) $p(n) = o\left(n^{1/2-\beta}\right)$ for some $\beta > 1/\alpha$, (II) $p(n)$ satisfies $\lim_{n\to\infty} n^{1/2} \sum_{j=p(n)+1}^{\infty} |\phi_j| = 0$.

Assumption **MU** is a uniform integrability condition, which is in general stronger than the one used by Chuang and Chan (2002) and Bravo (2005*b*) in the case of unstable autoregressive processes with martingale difference innovations. In this paper it is used to impose a restriction on the growth rate of $p(n)$, as specified in assumption **P**(I), which is the weaker the stronger **MU** is. Assumption **P**(II) is as in Berk (1974) and Said and Dickey (1984) and ensures the $n^{1/2}$ consistency of the estimates of the ϕ_j's[6]. At the same time **P**(II) rules out the logarithmic rate for $p(n)$, implying that information-based selection rules for $p(n)$, such as the Akaike information criterion (AIC) or the Bayesian (Schwartz) criterion (BIC) are not allowed. On the other hand **P**(II) allows using the sequential tests method of Ng and Perron (1995, p.270) for the choice of $p(n)$. This method seems to work well in practice and provides consistent estimates (the least squares estimates) that can be used as initial values for the ϕ_j's in the numerical algorithm used to compute $W\left(1,\widehat{\phi},\gamma\right)$.

Theorem 1 *Assume that M(I),(II) MU, LP and P(I),(II) hold. Then for any* $\gamma \in \Re$ *and* $p(n) = o\left(n^{1/4}\right)$

$$W\left(1,\widehat{\phi},\gamma\right) \Rightarrow \left(\int_0^1 B(r)\,dB(r)\right)^2 / \int_0^1 B^2(r)\,dr. \tag{14}$$

[6] It should be noted that **P**(II) is stronger than what is needed for the validity of this paper's results. Indeed Theorems 1 and 2 below are still valid if **P**(II) is replaced by the weaker assumption

P (II)' $p(n)$ satisfies $\lim_{n\to\infty} p(n)^{1/2} \sum_{j=p(n)+1}^{\infty} |\phi_j| = 0$.

Theorem 1 shows that $W\left(1, \widehat{\phi}, \gamma\right)$ converges to the square of the ADF t-statistic for a unit root defined in (3). Thus Theorem 1 provides an asymptotic justification for the test that rejects the unit root hypothesis at α level, when $W\left(1, \widehat{\phi}, \gamma\right) > u_{1-\alpha}$ where $u_{1-\alpha}$ is the $1-\alpha$ quantile of (14), which can be found by simulation. Note that the rate $o\left(n^{1/4}\right)$ is the same as that of Donald et al. (2003) in the context of consistent empirical likelihood tests for conditional moment restrictions.

To assess the power properties of $W\left(1, \widehat{\phi}, \gamma\right)$ we consider the same type of (local) alternative hypotheses considered by Chan and Wei (1987) and Phillips (1987*b*). It is important to note that in the context of empirical likelihood inference the distribution functions under the null and the local alternative hypotheses are related in the sense that they are both assumed to belong to the same class of multinomial distributions supported on the sample indexed by the parameter β. Bearing this in mind, the sequence of alternatives we consider is $H_n : \beta_n = 1 - \delta/n$ for some $-\infty < \delta < \infty$. Let $J_\delta(r) = \int_0^r \exp\left[-(r-s)\gamma\right] dB(s)$ denote an Ornstein-Uhlenbeck process.

Theorem 2 *Under the same assumptions of Theorem 1, for any $\gamma \in \Re$*

$$W\left(\beta_n, \widehat{\phi}, 1\right) \Rightarrow \left(\int_0^1 J_\delta(r)\, dB(r)\right)^2 / \int_0^1 J_\delta{}^2(r)\, dr. \tag{15}$$

Theorem 2 shows that $W\left(\beta_n, \widehat{\phi}, 1\right)$ has the same local asymptotic power of ADF_t^2. Since the rate of divergence under the alternative of the latter is $O_p(n)$ we may expect $W\left(\beta_n, \widehat{\phi}, 1\right)$ to have good power properties and in particular greater power than standard t-ratio test (11).

Remark 1. The profile SECR test statistic is a nondirectional test whereas the standard ADF t-ratio is directional against stationary or explosive alternatives. To obtain a directional profile SECR test statistic we can consider its squared root, that is

$$R\left(1, \widehat{\phi}, \gamma\right) := sign\left(\widehat{\beta} - 1\right)\left[W\left(1, \widehat{\phi}, \gamma\right)\right]^{1/2},$$

where $sign(\cdot)$ is the sign function.

Corollary 3 *Under the same assumptions of Theorem 1,*

$$R\left(1, \widehat{\phi}, \gamma\right) \Rightarrow \int_0^1 B(r)\, dB(r) / \left(\int_0^1 B^2(r)\, dr\right)^{1/2}.$$

Remark 2. Models with deterministic trends x_t can be analyzed similarly. Suppose that z_t is generated by

$$z_t = \eta' x_t + y_t \tag{16}$$

and y_t is given by (1). Then the unit root hypothesis can still be tested as in (2) using the residuals $\widehat{y}_t$ of a preliminary regression (16). It can be shown that the distribution of the resulting profile SERC statistic is

$$W_{\widehat{\eta}}\left(1, \widehat{\phi}, \gamma\right) \Rightarrow \left(\int_0^1 B_X(r)\, dB(r)\right)^2 / \int_0^1 B_X^2(r)\, dr$$

where $B_X(r) = B(r) - \left[\int_0^1 B(r) X(r)'\right] \left[\int_0^1 X(r) X(r)'\right]^{-1} X(r)$ is a detrended Brownian motion and depends on the limit trend function $X(r)$.

5 Sieve bootstrap for unit root tests

We now briefly consider the sieve bootstrap for ADF test as developed by Chang and Park (2003) (see Psaradakis (2001) for an alternative application of sieve bootstrap to unit root testing).

The sieve bootstrap procedure is motivated by the $AR(\infty)$ representation of the linear process u_t given by $\Phi(L) u_t = \varepsilon_t$ and its finite dimensional (sieve) $AR(p)$ approximation (9). Let $\widehat{\phi}_j$ $(j = 1, ..., p(n))$ denote the least squares estimates[7] of the autoregression (9) and let $\widehat{\varepsilon}_{p(n),t}$ denote the resulting residuals. As it is customary in bootstrap literature we use the asterisk $*$ to denote bootstrap samples. To obtain a bootstrap unit root process y_t^*, we first obtain i.i.d. samples $\varepsilon_{p(n),t}^*$ from the empirical distribution $\widehat{\varepsilon}_{p(n),t} - \sum_{t=1}^n \widehat{\varepsilon}_{p(n),t}/n$. Next we generate u_t^* from the fitted autoregression

$$u_t^* = \sum_{j=1}^{p(n)} \widehat{\phi}_j u_{t-j}^* + \varepsilon_{p(n),t}^* \tag{17}$$

with initial values $\begin{bmatrix} u_{1-p}^* & \dots & u_0^* \end{bmatrix}' = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}'$[8]. Finally using u_t^* the bootstrap

[7]Alternatively as suggested by Bühlmann (1997) and Psaradakis (2001) (8) can be estimated by the Yule -Walker method , which may be preferable in small samples since it always yields an invertible autoregression.

[8]Alternatively we may generate a large number of values of u_t^* and discard the first, say k, generated values to remove the effect of the initial values.

unit root process y_t^* is generated according to

$$y_t^* = y_{t-1}^* + u_t^* \quad t = 1, ..., n$$

with $y_0^* = 0$[9].

For the bootstrap version of the ADF test let

$$y_t^* = \beta y_{t-1}^* + \sum_{j=1}^{p(n)} \phi_j \Delta y_{t-j}^* + \varepsilon_{p(n),t}^*$$

and define the bootstrap sieve score function (cf. (10))

$$m_{p(n),t}^* (\beta, \phi) = \left[\begin{array}{cccc} y_{t-1}^* & \Delta y_{t-1}^* & \dots & \Delta y_{t-p(n)}^* \end{array} \right]' \left(y_t^* - \beta y_{t-1}^* - \sum_{j=1}^{p(n)} \phi_j \Delta y_{t-j}^* \right),$$

and let $\left[\begin{array}{cc} \widehat{\beta}^* & \widehat{\phi}'^* \end{array} \right]'$ denote the bootstrap least square estimator that solves

$$\sum_{t=1}^{n} m_{p(n),t}^* \left(\widehat{\beta}, \widehat{\phi} \right) = 0.$$

Then the bootstrap analogue of the ADF t-statistic for the unit root hypothesis $H_0 : \beta = 1$ is

$$ADF_t^* = \left(\widehat{\beta}^* - 1 \right) / \widehat{\sigma}_{\widehat{\beta}^*}^*$$

where $\widehat{\sigma}_{\widehat{\beta}^*}^*$ is the bootstrap standard error. Chang and Park (2003) show that the bootstrap ADF t-test is asymptotically valid. To be specific they show that for any $\alpha \in (0, 1)$

$$\Pr \left(ADF_t < u_\alpha^* | H_0 \right) \to \alpha$$

where $u_\alpha^* := \inf \left\{ u : \Pr^* \left(ADF_t^* \leq u \right) \geq \alpha \right\}$ is the bootstrap α-th quantile obtained by the bootstrap distribution of ADF_t^*.

The results of the previous section show that the SECR statistic is asymptotically equivalent to ADF_t^2. Thus to make the comparison between the SECR test statistic and the bootstrap test for a unit root meaningful we need to bootstrap the square of the ADF t- statistic. Let $u_{1-\alpha}^* := \inf \left\{ u : \Pr^* \left(ADF_t^{*2} \geq u \right) \geq 1 - \alpha \right\}$ and assume that the linear process (7) satisfies **LP**,

[9] As in the case of y_t the initial value of y_t^* does not affect the asymptotic results obtained in this section as long as $y_0^* = y_0 = O_p(1)$.

B The innovations ε_t are i.i.d. random variables with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$, $E(|\varepsilon_t|^4) < \infty$,[10]

and that the growth rate of the lag parameter $p(n)$ in (17) is

PB $p(n) = o\left((n/\log n)^{1/(2k+2)}\right)$[11] where k is defined in **LP**

Theorem 4 *Assume that* ***B****, and* ***PB****,* ***LP*** *holds for* $k = 1$*. Then*

$$Pr\left(ADF_t^2 > u_{1-\alpha}^* | H_0\right) \to 1 - \alpha \tag{18}$$

Remark 3. The growth rate $p(n) = o\left((n/\log n)^{1/4}\right)$ specified in Theorem 4 is slightly weaker than the corresponding one used in Theorems 1 and 2. Also assumption **PB** implies that the selection of the lag $p(n)$ in the sieve bootstrap approximation (17) can be based either on the sequential method as used in the previous section for the SECR statistic or on information-based selection rules.

Remark 4. As in the case of ECR statistic, the sieve bootstrap method extends easily to the case of autoregressive models with deterministic trends x_t as given in (16) using the detrended residuals $\widehat{y}_t$ instead of the original series z_t .

6 Monte Carlo evidence

In this section we use simulations to investigate the performance of the profile SECR test statistic for a unit root in finite samples. In the simulations we consider the two best known (and most used in practice) members of the SECR statistic -the empirical likelihood ratio $W\left(1, \widehat{\phi}, 0\right)$ and the Kullback-Liebler distance $W\left(1, \widehat{\phi}, -1\right)$- the squared ADF t-statistic ADF_t^2, and the bootstrap squared ADF t-statistic $BADF_t^2$ (cf. (18)).

[10]The i.i.d. assumption rather than the martingale difference of **M** and **MU** makes the usual bootstrap procedure meaningful.

[11]Assumption **PB** is used to prove the weak consistency (i.e. in probability) of the bootstrap ADF_t^2. To establish strong consistency of the bootstrap conditional distribution **PB** should be replaced with

PB' (I) $p(n) = o(n^{\kappa})$ with $0 < \kappa < 1/2$ (II) $p(n) = Cn^{\delta}$ for some constant C and $1/(4k) < \delta < 1/2$ (see (Chang and Park, 2003) for further details). Note that the lower bound condition **PB'**(II) is slightly stronger than **P**(II).

Remark 5. The choice of empirical likelihood ratio can also be motivated by the fact that it enjoys a number of interesting statistical properties (see Owen (2001) for more details), whereas that of Kullback-Liebler distance can be motivated by the numerical stability and robustness of the underlying nonparametric tilting method used in the estimation (see for example Imbens and Spady (2002)).

Remark 6. Note that the ADF_t^2 statistic corresponds to the so-called Euclidean likelihood (Owen, 1991), i.e. the ECR statistic $W\left(1, \widehat{\phi}, \gamma\right)$ with $\gamma = -2$.

The model we consider is

$$y_t = \beta y_{t-1} + u_t \quad t = 1, 2, ..., n$$

with $\beta = \{1, .95, .90\}$, and u_t is either $u_t = \theta u_{t-1} + \varepsilon_t$ or $u_t = \varepsilon_t + \theta \varepsilon_{t-1}$. In the simulations we use $\theta = \{-.8, -.5, -.2, 0, .2, .5, .8\}$ and three different specifications for the distribution of the error ε_t, namely $N(0,1)$ (standard normal), $\chi_4^2 - 4$ (centered chi-squared distribution with four degrees of freedom) and t_5 (t-distribution with five degrees of freedom). The sample size n is set to 100 and 200. All the samples are generated using the SPLUS functions `rnorm`, `rchi`, and `rt`.

A practical issue that arises in calculating $W\left(1, \widehat{\phi}, \gamma\right)$, ADF_t^2 and $BADF_t^2$ (or their square roots) is the choice of the lag $p(n)$ in the sieve approximation. There is a large body of simulation evidence showing that this choice has important implications for the finite sample properties of standard ADF tests for a unit root. We investigated this issue in a preliminary Monte Carlo study, where we considered four different specifications for $p(n) = \{2, 4, 6, 8\}$ as well as two different adaptive rules for choosing $p(n)$, one based on the sequential testing procedure suggested by Ng and Perron (1995) and the other based on the AIC criterion[12]. Table 1 reports the finite sample size of $W\left(1, \widehat{\phi}, -1\right)$, $W\left(1, \widehat{\phi}, 0\right)$, ADF_t^2 and $BADF_t^2$ in the case of a moving average specification of u_t with parameter values $\theta = \{-.8, .8\}$, $N(0,1)$ errors and sample size $n = 100$. Note that the empirical size is obtained from 10000 replications using simulated asymptotic critical values[13], whereas the bootstrap critical values are based

[12]Note however that under **P**(II) the AIC criterion is not allowed for any ECR statistics, so strictly speaking this criterion should not be considered.

[13]The critical values were obtained by approximating $B(r)$ with partial sums of $N(0,1)$ random variables with 5,000 steps and 99,999 replications.

on 1000 replications.

Table 1. Finite sample size of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$,
$W\left(1,\widehat{\phi},0\right) := W_0$, ADF_t^2 and $BADF_t^2$
for $MA\ N(0,1)$ errors at 0.05 nominal level

θ	$p(n)$	W_{-1}	W_0	ADF_t^2	$BADF_t^2$
	2	.432	.498	.448	.423
	5	.198	.223	.190	.153
−.8	8	.078	.089	.069	.054
	t_{seq}^1	.064	.069	.060	.053
	AIC	.187	.196	.144	.122
	2	.065	.070	.061	.055
	5	.075	.081	.064	.059
.8	8	.084	.092	.073	.064
	t_{seq}^1	.067	.068	.060	.053
	AIC	.073	.079	.078	.063

1 The lag $p(n)$ is selected from the range $2 \leq p(n) \leq 10$.

Table 1 indicates a number of interesting points: First the choice of $p(n)$ clearly has bearing on the size all the four test statistics considered. Second among various data dependent criteria for choosing $p(n)$ the one based on sequential testing suggested by Ng and Perron (1995) produced tests with better finite sample properties. Third contrary to the findings of Chang and Park (2003) the AIC criterion seems to produce bootstrap tests that are still characterized by some size distortion. Fourth empirical likelihood and Kullback-Liebler have very similar size properties.[14].

The results of Table 1 clearly suggest that the sequential testing procedure produces tests with the smallest size distortion. Accordingly the Monte Carlo evidence reported in the following tables is based on the lag $p(n)$ chosen from the range $2 \leq p(n) \leq p(n)_{\max}$ by means of a sequential 0.10 level two-sided t- test, with $p(n)_{\max} = 10$ for $n = 100$ and $p(n)_{\max} = 12$ for $n = 200$. The tables report the

[14]It should be noted however that in computing the empirical likelihood ratio we occasionally (about 5 per cent of the simulations) encountered numerical difficulties, whereas the computation of the Kullback-Liebler statistic was more reliable.

finite sample size and power of $W\left(1,\widehat{\phi},-1\right)$, $W\left(1,\widehat{\phi},0\right)$, ADF_t^2 and $BADF_t^2$ for the unit root hypothesis $H_0:\beta=1$ at the $\{0.10, 0.05\}$ nominal level. Note that the empirical size and power are obtained from 5000 and 1000 replications respectively, using the same simulated asymptotic critical values of Table 1, whereas the power is calculated using empirical critical values obtained from the simulations under the null hypothesis, and thus represents size-corrected power.

Tables 2-13 can be found after the Appendix.

The results of Tables 2-13 can be summarized as follows: First both Kullback-Liebler distance and empirical likelihood ratio statistic have size distortion similar to that of the square of ADF (euclidean likelihood) statistic, albeit slightly larger for $n=100$. Second Kullback-Liebler distance and empirical likelihood ratio have better size properties when the magnitude of the parameter θ is smaller. Third the bootstrapped square ADF statistic has in general the smallest size distortion. Fourth Kullback-Liebler distance and empirical likelihood ratio statistic have the best power properties, with the empirical likelihood having an edge over the Kullback-Liebler particularly when the alternative hypothesis is closer to the null hypothesis. Finally the bootstrapped square ADF statistic has noticeably less power than all of the other test statistics considered[15].

In practice these results suggest that if the errors are not strongly correlated there are some advantages (in terms of power) in using nonparametric likelihood based methods to test for a unit root (at least empirical likelihood or Kullback-Liebler). In particular for smaller sample sizes (e.g. $n\leq 100$) is small the Kullback-Liebler distance statistic is preferable to the empirical likelihood ratio on the grounds of its numerical stability. On the other hand if the errors are strongly correlated the bootstrap produces the most accurate tests for unit roots. However it is clearly not possible to know *a priori* what is the degree of correlation present in the errors. Thus from an applied point of view the following two-step procedure might be used: first fit a preliminary augmented (or sieve) regression as in (10) and use the sequential t test

[15]It should be noted that these results are not sample size dependent in the sense that we run the same Monte Carlo experiments for sample sizes $n=50$ and $n=400$ and obtained qualitatively the same type of results, with the only difference that for $n=50$ the empirical likelihood ratio was occasionally (i..e. about 10 per cent of simulations) numerically unstable.

method to choose the lag $p(n)$ specifying a large $p(n)_{\max}$. If the chosen $p(n)$ is such that $p(n) \leq p(n)_{\max}/2$[16] use the Kullback-Liebler distance or (or empirical likelihood ratio) statistics using as initial values the estimates obtained from the preliminary regression. Otherwise use the bootstrap with residuals obtained from the preliminary regression.

7 Conclusion

In this paper we have proposed a nonparametric likelihood based test statistic for a unit root in an autoregressive model with serially correlated errors. The test statistic is based on the ECR method and uses a sieve approximation to capture the dependent structure of the errors. We have derived the asymptotic distribution of the resulting sieve ECR statistic and assessed using simulations the finite sample properties of the two best known members: the empirical likelihood and the Kullback-Liebler distance. The results of the simulations suggest that both statistics have good finite sample properties, provided that the sample size is relatively large and the degree of dependency in the errors is small. These results suggest some directions for future research. First in the Monte Carlo simulation we have considered the empirical likelihood ratio and the Kullback-Liebler distance statistics. This choice was based on a number of statistical and numerical properties enjoyed by these two statistics, but clearly there are other test statistics that could be considered including the Hellinger (for $\gamma = -1/2$) and the Pearsonian Chi-square (for $\gamma = 1$) statistics .

Second the issue of selecting $p(n)$ requires further investigation. In the paper we have considered the method suggested by Ng and Perron (1995). This method seems to work well in practice, however it needs not to be optimal. Also the proposed two-step procedure based on the crude upper bound $p(n)_{\max}/2$ could be improved.

Third the sieve bootstrap unit root tests are very accurate (i.e. their empirical size is very close to the nominal size) but they might lack in power. This suggests that using a method that combines the bootstrap and the ECR (or any other nonparametric likelihood methods) approach, like for example the so-called *biased* bootstrap

[16] It may appear that this choice is somewhat arbitrary. However it does seem to work in practice -at least for the ARMA type of errors that we considered in this paper and in other simulations experiment not reported here.

(Hall and Presnell, 1999), might result in unit roots test statistics with remarkably good finite sample properties both in terms of size and of power. This possibility is certainly of interest and is currently under investigation.

References

Baggerly, K. A. (1998), 'Empirical likelihood as a goodness of fit measure', *Biometrika* **85**, 535–547.

Basawa, I., Mallikand, A., McCormick, W., Reeves, J. and Taylor, L. (1991), 'Bootstrapping unstable first-order autoregressive processes', *Annals of Statistics* **19**, 1098–1101.

Berk, N. (1974), 'Consistent autoregressive spectral estimates', *Annals of Statistics* **2**, 489–502.

Beveridge, S. and Nelson, C. (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle"', *Journal of Monetary Economics* **7**, 151–174.

Bravo, F. (2005*a*), 'Bartlett type adjustments for empirical discrepancy test statistics', *Journal of Statistical Planning and Inference* . Forthcoming.

Bravo, F. (2005*b*), Nonparametric likelihood inference for general autoregressions. Submitted for publication.

Bravo, F. (2005*c*), Sieve empirical likelihood for unit root tests. Submitted for publication.

Bühlmann, P. (1997), 'Sieve bootstrap for time series', *Bernoulli* **3**, 123–148.

Chan, N. and Wei, C. (1987), 'Asymptotic inference of nearly nonstationary AR(1) process', *Annals of Statistics* **15**, 1050–1063.

Chan, N. and Wei, C. (1988), 'Limiting distributions of least squares estimates of unstable autoregressive processes', *Annals of Statistics* **16**, 367–401.

Chang, Y. and Park, J. (2002), 'On the asymptotics of ADF tests for unit roots', *Econometric Reviews* **21**, 449–476.

Chang, Y. and Park, J. (2003), 'A sieve bootstrap for the test of a unit root', *Journal of Time Series Analysis* **24**, 379–400.

Chuang, C. and Chan, N. (2002), 'Empirical likelihood for autoregressive models, with applications to unstable time series', *Statistica Sinica* **12**, 387–408.

Corcoran, S. A. (1998), 'Bartlett adjustments of empirical discrepancy statistics', *Biometrika* **85**, 967–972.

DiCiccio, T. and Romano, J. P. (1990), 'Nonparametric confidence limits by resampling methods and least favourable families', *International Statistical Review* **58**, 59–76.

Dickey, D. and Fuller, W. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**, 427–431.

Donald, S., Imbens, G. and Newey, W. (2003), 'Empirical likelihood estimation and consistent tests with conditional moment restrictions', *Journal of Econometrics* **117**, 55–93.

Efron, B. (1981), 'Nonparametric standard errors and confidence intervals (with discussion)', *Canadian Journal of Statistics,* **9**, 139–172.

Hall, P. and Heyde, C. (1980), *Martingale Limit Theory and its Applications*, Academic Press: New York.

Hall, P. and Presnell, B. (1999), 'Intentionally biased bootstrap methods', *Journal of the Royal Statistical Society B* **61**, 143–158.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.

Imbens, G. and Spady, R. (2002), 'Confidence intervals in generalized method of moments models', *Journal of Econometrics* **107**, 87–98.

Lai, T. and Wei, C. (1982), 'Least squares estimates in stochastic regression models with applications to identification and control of dynamic system', *Annals of Statistics* **10**, 154–166.

Lee, S. and Young, G. (1999), 'Nonparametric likelihood ratio confidence inetrvals', *Biometrika* **86**, 107–118.

Newey, W. and Smith, R. (2004), 'Higher order properties of GMM and generalized empirical likelihood estimators', *Econometrica* **72**, 219–256.

Ng, S. and Perron, P. (1995), 'Unit root tests in ARMA models with data-depandent methods for the selection of the truncation lag', *Journal of the American Statistical Association* **90**, 268–281.

Owen, A. (1988), 'Empirical likelihood ratio confidence intervals for a single functional', *Biometrika* **36**, 237–249.

Owen, A. (1991), 'Empirical likelihood for linear models', *Annals of Statistics* **19**, 1725–1747.

Owen, A. (2001), *Empirical Likelihood*, Chapman and Hall.

Park, J. (2002), 'An invariance principle for sieve bootstrap in time series', *Econometric Theory* **18**, 469–490.

Phillips, P. (1987*a*), 'Time series regression with a unit root', *Econometrica* **55**, 277–301.

Phillips, P. (1987*b*), 'Towards a unified asymptotic theory for autoregression', *Biometrika* **74**, 535–547.

Phillips, P. and Pierron, P. (1988), 'Testing for a unit root in time series regressions', *Biometrika* **75**, 335–346.

Phillips, P. and Xiao, Z. (1998), 'A primer on unit root testing', *Journal of Economics Surveys* **12**, 423–470.

Psaradakis, Z. (2001), 'Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors', *Journal of Time Series Analysis* **22**, 577–594.

Read, T. R. C. and Cressie, N. (1988), *Goodness-of-Fit Statistics for Discrete Multivariate Data*, New York: Springer-Verlag.

Romano, J. and Wolf, M. (2001), 'Subsampling intervals in autoregressive models with linear time trend', *Econometrica* **69**, 1283–1314.

Said, S. and Dickey, D. (1984), 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika* **71**, 599–608.

Smith, R. (1997), 'Alternative semi-parametric likelihood approaches to generalised method of moments estimation', *Economic Journal* **107**, 503–519.

Xiao, Z. and Phillps, P. (1998), 'An ADF coefficient test for a unit root in ARMA models with unknown order with empirical applications to the US economy', *Econometrics Journal* **1**, 27–43.

Appendix

7.1 Duality between ECR and generalized empirical likelihood

Let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval V containing 0 with derivatives $\rho_j(v) = d^2\rho(v)/dv^j$ $(j = 1, 2, ...)$ and let $V_n(\theta) = \{\lambda : \lambda' g_i(x_i, \theta_0) \in V,\ i = 1, ...n\}$. A generalized empirical likelihood (GEL) statistic (Smith, 1997) for $H_0 : \theta = \theta_0$ in (1) is given by

$$2\sum_{i=1}^{n}\left(\rho\left(\widehat{\lambda}' g_i(x_i, \theta_0)\right)/n - \rho(0)\right)$$

where $\widehat{\lambda} := \arg\max_{\lambda \in V_n(\theta_0)} \sum_{i=1}^{n} \rho\left(\widehat{\lambda}' g_i(x_i, \theta_0)\right)/n$. Associated with any GEL statistic are the so-called implied probabilities $\widehat{\pi}_i = \rho_1\left(\widehat{\lambda}' g_i(x_i, \theta_0)\right)/\sum_{i=1}^{n}\rho_1\left(\widehat{\lambda}' g_i(x_i, \theta_0)\right)$ which by construction sum to 1, and satisfy the sample moment condition

$$\sum_{i=1}^{n}\widehat{\pi}_i g_i(x_i, \theta_0) = 0$$

when the first order conditions for λ hold, mirroring the population moment condition (1).

The solution (4) to the constrained optimization problem (3) can be written for $\delta \neq 0$ as

$$\widetilde{w}_i(\theta_0, \delta) = \left(1 + \delta\widehat{\eta}' g(x_i, \theta_0)\right)^{-1/\delta} / \sum_{i=1}^{n}\left(1 + \delta\widehat{\eta}' g(x_i, \theta_0)\right)^{-1/\delta} \tag{19}$$

where $\delta = 1 + \gamma$ and $\widehat{\eta} = \widehat{\xi}/\left(\delta\widehat{\zeta}\right)$. By defining

$$\rho(v) = (1 + \delta v)^{(\delta-1)/\delta}/(\delta - 1) \tag{20}$$

it is easy to see that $\rho_1(v) = (1 + \delta v)^{-1/\delta}$ and therefore the implied probabilities are

$$\widehat{\pi}_i = \left(1 + \delta\widehat{\lambda}' g_i(x_i, \theta_0)\right)^{-1/\delta} / \sum_{i=1}^{n}\left(1 + \delta\widehat{\lambda}' g_i(x_i, \theta_0)\right)^{-1/\delta}. \tag{21}$$

Comparing (19) and (21) clearly shows that for $\widehat{\lambda} = \widehat{\eta}$ there is a dual relationship between ECR and GEL method, that is every ECR statistic can be obtained as a

GEL statistic and viceversa every GEL statistic based on (20) can be interpreted as an ECR statistic. This duality, first noted by Newey and Smith (2004), is useful in the context of this paper because it implies that the asymptotic properties of ECR statistic can be recovered from those of the corresponding GEL statistic, which are somewhat simpler to derive.

7.2 Notation

In what follows C denotes a generic positive constant, $\sum = \sum_{t=1}^{n}$, unless otherwise stated, and "CMT" stands for continuous mapping theorem. Also if M is a symmetric matrix and v is a vector the matrix norm $\|M\|$ is defined as $\sup_{(v'v)^{1/2}\leq 1}\|Mv\|$, whereas $\varsigma_{\max}(M)$ and $\varsigma_{\min}(M)$ denote the largest and smallest eigenvalue of M. For notational convenience, let $p(n) := p$ and let

$$\begin{aligned}
\Delta\mathbf{y}_{p,t} &= \begin{bmatrix} \Delta y_t & \dots & \Delta y_{t-p+1} \end{bmatrix}', \quad S_n = diag\begin{bmatrix} n^{-1} & n^{-1/2}I_p \end{bmatrix}, \\
Y_{t-1,n} &= S_n \begin{bmatrix} y_{t-1} & \Delta\mathbf{y}'_{p,t-1} \end{bmatrix}', \quad m_{tn} = Y_{t-1,n}\varepsilon_{p,t}, \\
M_{nn} &= E\left(\sum m_{tn}m'_{tn}|\mathcal{F}_{t-1}\right), \quad \Sigma_{\omega p} = diag\left(\sigma^2\begin{bmatrix} \omega & \Sigma_p \end{bmatrix}\right) \\
m_{tn}(\phi) &= Y_{t-1,n}\left(y_t - y_{t-1} - \Delta\mathbf{y}'_{p,t}\phi\right), \quad \widetilde{m}_t := m_{tn}\left(\widetilde{\phi}\right) \\
\Phi_{\tau_{pn}} &= \|\phi - \phi_0\| \leq \tau_n(p) := \tau_{pn} \text{ and } \tau_{pn}p \to 0 \text{ as } n \to \infty
\end{aligned}$$

where $\varepsilon_{p,t} = \varepsilon_t + \sum_{j=p+1}^{\infty}\phi_j u_{t-j}$, ω is as random variable such that $\Pr(\omega > 0) \to 1$, and the $\Re^{p\times p}$ valued matrix $\Sigma_p = O(1)$ has typical (i,j) element $E(\Delta y_{t-i}\Delta y_{t-j})$.

7.3 Preliminary lemmae

Lemma 5 *Assume that* **M** *and* **LP** *hold. Then for* $p = o\left(n^{1/2}\right)$

$$\|M_{nn} - \Sigma_{\omega p}\| = o_p(1). \tag{22}$$

Proof. Note that $E\left[(\varepsilon_{p,t} - \varepsilon_t)^2|\mathcal{F}_{t-1}\right] \leq E\left(u_t^2|\mathcal{F}_{t-1}\right)\left(\sum_{j=p+1}^{\infty}|\phi|\right)^2 = o_{a.s.}(1)$, so by triangle inequality $E\left(\varepsilon_{p,t}^2|\mathcal{F}_{t-1}\right) = \sigma^2 + o_{a.s.}(1)$, which implies that $E\left(\sum m_{tn}m'_{tn}|\mathcal{F}_{t-1}\right) = \sigma^2\sum Y_{t-1,n}Y'_{t-1,n} + o_{a.s.}(1)$. By Berk (1974) $\sum\Delta\mathbf{y}_{p,t-1}\Delta\mathbf{y}'_{p,t-1}/n \xrightarrow{p} \Sigma_p$ and by Phillips (1987*b*) $\sum y_{t-1}^2/n^2 \xrightarrow{w} \sigma^2\Psi(1)\int B^2(r)\,dr := \sigma^2\omega$. Furthermore, since $\left\|\sum y_{t-1}\Delta\mathbf{y}'_{p,t-1}/n\right\| = O_p\left(p^{1/2}\right)$ (Chang and Park, 2002, Lemma 3.2), $\sum y_{t-1}\Delta\mathbf{y}'_{p,t-1}/n^{3/2} \xrightarrow{p} 0$. Thus by iterated expectations it can be shown that $E\|M_{nn}-\Sigma_{\omega p}\|^2 = Cp^2/n$ which implies (22). ∎

Lemma 6 *Assume that* **M** *and* **LP** *hold. Then for $p = o\left(n^{1/2}\right)$ and $\eta \to \infty$*

$$\sup_n \Pr\{\|M_{nn}\| > \eta\} \to 0. \tag{23}$$

Proof. Note that $\|M_{nn}\| \le \varsigma_{\max}(M_{nn})$ and that by (22) $|\varsigma_{\max}(M_{nn}) - \varsigma_{\max}(\Sigma_{\omega p})| \le \|M_{nn}-\Sigma_{\omega p}\| = o_p(1)$. Since $\varsigma(\Sigma_{\omega p}) = O_p(1)$, it then follows that $\varsigma(M_{nn}) = O_p(1)$ whence (23). ∎

Lemma 7 *Assume that* **M** *and* **LP** *hold. Then for $p = o\left(n^{1/2}\right)$ and any $\epsilon > 0$*

$$\sum E\left(\|m_{tn}\|^2 I\{\|m_{tn}\| > \epsilon\} |\mathcal{F}_{t-1}\right) = o_p(1). \tag{24}$$

Proof. Note that

$$\begin{aligned}\sum E\left(\|m_{tn}\|^2 I\{\|m_{tn}\| > \epsilon\} |\mathcal{F}_{t-1}\right) &\le \sum \|Y_{t-1,n}\|^4 E\left(|\varepsilon_{p,t}|^4 |\mathcal{F}_{t-1}\right)/\epsilon^2 \\ &\le CO_p\left(p^2/n\right) = o_p(1)\end{aligned}$$

since $y_{t-1}/n^{1/2} = O_p(1)$ (Chan and Wei, 1988). ∎

Lemma 8 *Assume that* **M** *and* **LP** *hold. Then for $p = o\left(n^{1/2}\right)$*

$$\left\|\sum m_{tn}m'_{tn} - \Sigma_{\omega p}\right\| = o_p(1) \tag{25}$$

and

$$\left|\varsigma\left(\sum m_{tn}m'_{tn}\right) - \varsigma(\Sigma_{\omega p})\right| = o_p(1) \tag{26}$$

where $\varsigma(\cdot)$ is either $\varsigma_{\min}(\cdot)$ or $\varsigma_{\max}(\cdot)$.

Proof. Note that (23) and (24) imply Theorem 2.23 of Hall and Heyde (1980) and thus

$$\left\|\sum m_{tn}m'_{tn} - M_{nn}\right\| = o_p(1).$$

Therefore (25) follows by triangle inequality, whereas (26) follows by (25). ∎

Lemma 9 *Let $y_t = y_{t-1} + u_t$ where $u_t = \Psi(L)\varepsilon_t = \sum_{j=1}^{\infty}\psi_j\varepsilon_{t-j}$, ε_t is a martingale difference sequence satisfying* ***M****, $\Psi(L)$ satisfies* ***LP*** *and $y_0 = 0$. Then $\max_t |y_t/n| = O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right)$ and $\max_t |u_t| = o_{a.s.}(n^{\beta})$ for $\beta > 1/\alpha$*

Proof. By the Beveridge-Nelson decomposition (Beveridge and Nelson, 1981) $y_t = \Psi(1)S_t + \eta_t$ where $S_t = \sum_{j=1}^{t}\varepsilon_j$, $\eta_t = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$ and $\alpha_j = -\sum_{k=1}^{\infty}\psi_{j+k}$. By Lai and Wei (1982) $\max_t |S_t| = O_{a.s.}\left(n^{1/2}(\log\log n)^{1/2}\right)$ and $\max_t |\varepsilon_t| = o_{a.s.}(n^{\beta})$, so that the conclusion follows by CMT noting that $\max_t |u_t| \leq \sum_{j=}^{\infty}|\psi_j|\max_t|\varepsilon_t| = o_{a.s.}(n^{\beta})$. ■

Lemma 10 *Assume that* ***M*** *and* ***LP*** *hold. Let $R_n = o\left(diag\begin{pmatrix} n^{-1} & n^{-\beta}I_p \end{pmatrix}\right)$ for $\beta > 1/\alpha$, $\Lambda_{pn} = \left\{\lambda : p^{1/2}\left\|R_n^{-1}\lambda\right\| \leq \epsilon\right\}$ where $\epsilon \leq 1$. Then*

$$\max_t \sup_{\lambda\in\Lambda_{pn}} |\lambda' m_t| = o_{a.s.}(1), \tag{27}$$

and

$$\Lambda_{pn} \subset \Lambda_0 \quad a.s. \tag{28}$$

where Λ_0 is an open interval containing 0.

Proof. By Lemma 9 $\max_t |y_{t-1}/n| = o_{a.s.}(1)$ and $\max_{j,t}|\Delta y_{t-j}| = o_{a.s}(n^{\beta})$ for $j = 1, ..., p$ so that $\max_t \|R_n m_t\| = o_{a.s.}(p^{1/2})$. Then on Λ_{pn} $\max_t \sup_{\lambda\in\Lambda_{pn}}|\lambda' m_t| \leq \epsilon/p^{1/2}\max_t\|R_n m_t\| = o_{a.s.}(1)$ which implies (28). ■

Lemma 11 *Assume that* ***M(I),(II), MU, LP, P(I),(II)*** *hold, and let*

$$\widehat{\lambda} := \arg\max_{\lambda\in\Lambda_0}\sum\rho(\lambda' m_t),$$

where $\rho(\cdot)$ is as in (20). Then $\widehat{\lambda}$ exists a.s., $\left\|S_n^{-1}\widehat{\lambda}\right\| = O_p(p^{1/2})$ and

$$\sup_{\lambda\in\Lambda_0}\sum\rho(\lambda' m_t) \leq \rho(0) + O_p(p). \tag{29}$$

Proof. On Λ_{pn} $\sum\rho(\lambda' m_t)$ is twice continuously differentiable so that $\widehat{\lambda} := \arg\max_{\Lambda_{pn}}\sum\rho(\lambda' m_t)$ exists $a.s.$ Let $\widehat{\lambda}_n = \kappa_n\theta$ where $\kappa_n = \left\|\widehat{\lambda}_n\right\|$, $\widehat{\lambda}_n = S_n^{-1}\widehat{\lambda}$ and

$\|\theta\| = 1$. Lemma 10 implies that for $0 \leq \overline{\lambda} \leq \widehat{\lambda}$ $\max_t \rho_2\left(\overline{\lambda}_n' m_{tn}\right) \leq -C$, thus by Taylor expansion

$$\begin{aligned}
\rho(0) &\leq \sum \rho\left(\widehat{\lambda}_n' m_{tn}\right) = \rho(0) + \widehat{\lambda}_n' \sum m_{tn} + \sum \rho_2\left(\overline{\lambda}_n' m_{tn}\right) \widehat{\lambda}_n' m_{tn} m_{tn}' \widehat{\lambda}_n / 2 \\
&\leq \rho(0) + \kappa_n \left\| \sum m_{tn} \right\| - C\kappa_n^2 \theta' \sum m_{tn} m_{tn}' \theta \\
&\leq \rho(0) + \kappa_n \left\| \sum m_{tn} \right\| - C\kappa_n^2 \qquad (30)
\end{aligned}$$

where the last inequality follows because $\theta' \sum m_{tn} m_{tn}' \theta \geq \varsigma_{\min}(\Sigma_{\omega p}) > 0$ *a.s.*by (26). Subtracting $\rho(0) - C\kappa_n^2$ from both sides and dividing by κ_n we find that

$$\kappa_n \leq \left\| \sum m_{tn} \right\| = O_p\left(p^{1/2}\right)$$

where the last equality follows by

$$\begin{aligned}
&E\left\| \sum \Delta \mathbf{y}_{p,t-1}' \varepsilon_{p,t}/n \right\|^2 \leq E\left\| \sum \Delta \mathbf{y}_{p,t-1}'/n \right\|^2 E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right) + \\
&E\left\| \sum \Delta \mathbf{y}_{p,t-1}' \right\|^2 E\left(\left(\varepsilon_{p,t} - \varepsilon_t\right)^2 | \mathcal{F}_{t-1}\right) \leq O(p) + o\left(p^{1-2k}\right) \to 0.
\end{aligned}$$

Note that for any $\delta_n = O(1)$ $\left\| S_n^{-1} \widehat{\lambda} \right\| \leq \delta_n p^{1/2}$ implies that

$$p^{1/2} \left\| R_n^{-1} \widehat{\lambda} \right\| \leq \varsigma_{\max}\left(S_n R_n^{-1}\right) \delta_n p = O\left(n^{-1/2+\beta}\right) p \to 0$$

for $p = o\left(n^{1/2-\beta}\right)$. Thus $p^{1/2} \left\| S_n^{-1} \widehat{\lambda} \right\| < \epsilon$ *a.s.* that is $\left\| \widehat{\lambda} \right\| \in int\{\Lambda_{pn}\}$ and hence $\left\| \widehat{\lambda} \right\| \in int\{\Lambda_0\}$. By concavity of $\sum \rho(\lambda' m_t)$ and convexity of Λ_0 it then follows that $\widehat{\lambda} := \arg\max_{\Lambda_0} \sum \rho(\lambda' m_t)$ exists *a.s.* Finally (29) follows by (30) because $\kappa_n = O_p\left(p^{1/2}\right)$. ■

Lemma 12 *Let* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$. *Then*

$$\left\| \sum \widetilde{m}_{tn} \widetilde{m}_{tn}' - \sum m_{tn} m_{tn}' \right\| = o_p(1). \qquad (31)$$

Proof. By triangle and Cauchy-Schwartz inequalities

$$\begin{aligned}
&\left\| \sum \widetilde{m}_{tn} \widetilde{m}_{tn}' - \sum m_{tn} m_{tn}' \right\| \leq \sum \left| \widetilde{\varepsilon}_{t,p}^2 - \varepsilon_{t,p}^2 \right| \left\| Y_{t-1,n} \right\|^2 \\
\leq &\sum \left| \left(\widetilde{\varepsilon}_{t,p} - \varepsilon_{t,p}\right)^2 - \left(\widetilde{\varepsilon}_{t,p} - \varepsilon_{t,p}\right) \varepsilon_{t,p} \right| \left\| Y_{t-1,n} \right\|^2 \\
\leq &C \left\| \widetilde{\phi} - \phi_0 \right\| \sum \left\| Y_{t-1,n} \right\|^2 = O_p\left(\tau_{pn} p\right) \to 0.
\end{aligned}$$

■

Lemma 13 *Assume that* ***M(I),(II), MU,*** *and* ***P(I)*** *hold. Then on the set* Λ_{pn} *as defined in Lemma 10 for any* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$

$$\max_t \sup_{\widetilde{\phi}\in\Phi_{pn}} \sup_{\lambda\in\Lambda_{pn}} \left|\lambda'\widetilde{m}_t\right| = o_p(1), \quad \textit{and } \Lambda_{pn} \subset \Lambda_0(\phi) \quad a.s. \tag{32}$$

where $\Lambda_0(\phi)$ *is an open interval (depending on* ϕ*) containing 0.*

Proof. Note that $\widetilde{m}_t = \begin{bmatrix} y_{t-1} & \Delta\mathbf{y}'_{p,t-1} \end{bmatrix}' \left(\varepsilon_{t,p} - \Delta\mathbf{y}'_{p,t-1}\left(\widetilde{\phi}-\phi_0\right)\right)$. By Lemma 9 $\max_t \left\|\Delta\mathbf{y}_{p,t-1}\right\| = o_{a.s.}\left(p^{1/2}n^{\beta}\right)$ and thus by triangle inequality

$$\begin{aligned} &\max_t \sup_{\widetilde{\phi}\in\Phi_{pn}} \sup_{\lambda\in\Lambda_{pn}} \left|\lambda'\widetilde{m}_t\right| \leq \epsilon/p^{1/2} \max_t \left(\left\|R_n m_t\right\| + \left\|R_n \begin{bmatrix} y_{t-1} & \Delta\mathbf{y}'_{p,t-1}\end{bmatrix}' \Delta\mathbf{y}'_{p,t-1}\right\| \tau_{pn}\right) \\ = \;& \epsilon/p^{1/2}\left(o_{a.s.}\left(p^{1/2}\right) + o_{a.s.}\left(\tau_{pn}p\right)\right), \end{aligned}$$

so that (32) follows as in Lemma 10. ■

Lemma 14 *Assume that* ***M(I),(II), MU, LP, P(I),(II)*** *hold. Let* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$*,*

$$\widehat{\lambda} := \arg\max_{\lambda\in\Lambda_0(\widetilde{\phi})} \sum \rho\left(\lambda'\widetilde{m}_t\right)$$

and suppose that $\left\|\widetilde{m}_{tn}\right\| = O_p\left(p^{1/2}\right)$*. Then* $\widehat{\lambda}$ *exists a.s.,* $\left\|S_n^{-1}\widehat{\lambda}\right\| = O_p\left(p^{1/2}\right)$

$$\sup_{\lambda\in\Lambda_0(\widetilde{\phi})} \sum \rho\left(\lambda'\widetilde{m}_t\right) \leq \rho(0) + O_p(p). \tag{33}$$

Proof. As in the proof of Lemma 11 $\sum\rho\left(\lambda'\widetilde{m}_t\right)$ is twice differentiable on Λ_{pn} and thus $\widehat{\lambda} := \arg\max_{\lambda\in\Lambda_{pn}} \sum\rho\left(\lambda'\widetilde{m}_t\right)$ exists *a.s.* Using the same argument and notation of Lemma 11 we get

$$0 \leq \kappa_n \left\|\sum \widetilde{m}_{tn}\right\| - C\kappa_n^2\theta' \sum \widetilde{m}_{tn}\widetilde{m}'_{tn}\theta.$$

Then by (26), (31), (32) and CMT $\varsigma_{\max}\left(\sum \widetilde{m}_{tn}\widetilde{m}'_{tn}\right) \leq C$ and $\max_t \left|\lambda'_n\widetilde{m}_{tn}\right| = o_p(1)$, and thus

$$\kappa_n \leq \left\|\sum \widetilde{m}_{tn}\right\| = O_p\left(p^{1/2}\right),$$

so that $\left\|S_n^{-1}\widehat{\lambda}\right\| = O_p\left(p^{1/2}\right)$, and (33) follows as in Lemma 11. ■

Lemma 15 *Assume that* ***M(I),(II), MU, LP, P(I),(II)*** *hold. Let* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$ *and let* $\widehat{\phi} := \arg\min_{\widetilde{\phi}\in\Phi_{\tau_{pn}}} \sum\rho\left(\lambda'\widetilde{m}_t\right)$*. Then* $\left\|\widehat{m}_{tn}\right\| = O_p\left(p^{1/2}\right)$*.*

Proof. Let $\overline{\lambda} = R_n \widehat{m}_{tn} / \|\widehat{m}_{tn}\| p^{1/2}$ so that $\overline{\lambda} \in \Lambda_{np}$ and hence $\max_t \left|\overline{\lambda}' \widehat{m}_t\right| = o_{a.s.}(1)$ by (32). Note that $\max_t -\rho_2\left(\overline{\lambda}' \widehat{m}_{tn}\right) \leq C$ *a.s.* so that

$$\sum \rho_2\left(\overline{\lambda}' \widehat{m}_{tn}\right) \widehat{m}_{tn} \widehat{m}_{tn}' \geq C \sum \widehat{m}_{tn} \widehat{m}_{tn}' \ a.s.$$

and $\varsigma_{\min}\left(\sum \widehat{m}_{tn} \widehat{m}_{tn}'\right) \geq C$ by (31). Thus similarly to Lemma 11 for $0 \leq \widetilde{\lambda} \leq \overline{\lambda}$

$$\begin{aligned} \sum \rho\left(\overline{\lambda}' \widehat{m}_t\right) &= \overline{\lambda}' S_n^{-1} \widehat{m}_{tn} + \sum \rho_2\left(\widetilde{\lambda}_n' m_{tn}\right) \overline{\lambda}' S_n^{-1} \widehat{m}_{tn} \widehat{m}_{tn}' S_n^{-1} \overline{\lambda}/2 \\ &\geq \overline{\lambda}' S_n^{-1} \widehat{m}_{tn} - C \overline{\lambda}' S_n^{-1} S_n^{-1} \overline{\lambda} \\ &= \varsigma_{\max}\left(R_n S_n^{-1}\right) \|\widehat{m}_{tn}\| / p^{1/2} - C \varsigma_{\max}\left(R_n S_n^{-1}\right)^2 / p. \end{aligned}$$

Note that by the definition of $\widehat{\phi}$ and (33)

$$\sup_{\lambda \in \Lambda_{pn}(\widehat{\phi})} \sum \rho\left(\lambda' \widehat{m}_t\right) \leq \sup_{\lambda \in \Lambda_{pn}} \sum \rho\left(\lambda' m_t\right) = O_p(p)$$

and thus

$$O_p(p) \geq \sum \rho\left(\overline{\lambda}' \widehat{m}_t\right) \geq \varsigma_{\max}\left(R_n S_n^{-1}\right) \|\widehat{m}_{tn}\| / p^{1/2} - C O_p\left(\varsigma_{\max}\left(R_n S_n^{-1}\right)^2\right) / p$$

which implies that $\|\widehat{m}_{tn}\| \leq O_p\left(p^{3/2} / n^{1/2-\beta}\right) + O_p\left(n^{1/2-\beta} / p^{1/2}\right) = o_p\left(p^{1/2}\right) + O_p\left(p^{1/2}\right)$.
■

Lemma 16 *Assume that* **M(I),(II), MU, LP, P(I),(II)** *hold. Let $\widetilde{\phi} \in \Phi_{\tau_{pn}}$ and suppose that $\tau_{pn} = o\left(1/\left(n^{\beta} p^{1/2}\right)\right)$. Then there exists a $\widehat{\phi} \in int\left\{\Phi_{\tau_{pn}}\right\}$ minimizing $\sum \rho\left(\lambda' \widetilde{m}_t\right)$ a.s. satisfying $0 = \sum \partial\left[\rho\left(\lambda' \widetilde{m}_t\right)\right] / \partial \phi$.*

Proof. By a Taylor expansion we have

$$\begin{aligned} \sum \rho\left(\lambda' \widetilde{m}_t\right) &= \sum \rho\left(\lambda' m_t\right) + \sum \partial \rho\left(\lambda' m_t\right) / \partial \phi\left(\widetilde{\phi} - \phi_0\right) + \\ & \left(\widetilde{\phi} - \phi_0\right)' \sum \partial^2 \rho\left(\lambda' m_t\right) / \partial \phi \partial \phi' \left(\widetilde{\phi} - \phi_0\right) / 2 + \breve{\xi}_n \end{aligned} \quad (34)$$

where

$$\breve{\xi}_n = \left(\widetilde{\phi} - \phi_0\right)' \sum \left[\partial^2 \rho\left(\lambda' \breve{m}_t\right) / \partial \phi \partial \phi' - \partial^2 \rho\left(\lambda' m_t\right) / \partial \phi \partial \phi'\right] \left(\widetilde{\phi} - \phi_0\right) / 2$$

and $\breve{m}_t$ is evaluated at the mean value $\breve{\phi}$ between $\widetilde{\phi}$ and ϕ_0. Lemma 15 implies that as long as $\widetilde{\phi} \in \Phi_{\tau_{pn}}$ $\|\widetilde{m}_{tn}\| = O_p\left(p^{1/2}\right)$ and thus $\|\breve{m}_{tn}\| = O_p\left(p^{1/2}\right)$. Also (31) and a Taylor expansion show that on Λ_{pn}

$$\lambda_n = \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum m_{tn}. \quad (35)$$

Let

$$\begin{aligned} \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}' &: = P_n^2, \\ \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum m_{tn} &: = Q_n; \end{aligned}$$

by the chain rule for $1 \leq j_1, ... j_k \leq p$

$$\partial^k \rho \left(\lambda' m_t\right) / \partial \phi_{j_1} ... \partial \phi_{j_k} = \left[\partial^k \rho \left(\lambda' m_t\right) / \partial \left(\lambda' m_t\right)^k\right] \partial \left(\lambda' m_t\right)^k / \partial \phi_{j_1} ... \partial \phi_{j_k} \tag{36}$$

so that using (36) for $k = 1$ and 2, and (35) gives

$$\begin{aligned} \sum \partial \rho \left(\lambda' m_t\right) / \partial \phi &= -\rho_1 \left(\lambda' m_t\right) Q_n \\ \sum \partial^2 \rho \left(\lambda' m_t\right) / \partial \phi \partial \phi' &= \rho_2 \left(\lambda' m_t\right) P_n^2. \end{aligned} \tag{37}$$

Inserting (37) into (34) yields

$$\begin{aligned} \sum \rho \left(\lambda' \widetilde{m}_t\right) &= \sum \rho \left(\lambda' m_t\right) - \rho_1 \left(\lambda' m_t\right) Q_n \left(\widetilde{\phi} - \phi_0\right) + \\ & \left(\widetilde{\phi} - \phi_0\right)' \rho_2 \left(\lambda' m_t\right) P_n^2 \left(\widetilde{\phi} - \phi_0\right) / 2 + \breve{\xi}_n \end{aligned} \tag{38}$$

where

$$\breve{\xi}_n = \left(\widetilde{\phi} - \phi_0\right)' \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left[\left(\sum \rho_2 \left(\lambda' \breve{m}_t\right) \breve{m}_{tn} \breve{m}_{tn}'\right)^{-1} - \left(\sum \rho_2 \left(\lambda' m_t\right) m_{tn} m_{tn}'\right)^{-1}\right] \times \sum Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}' \left(\widetilde{\phi} - \phi_0\right) / 2$$

By (32) $\max_t \left|\rho_k \left(\lambda' m_t\right) - \rho_k (0)\right| = o_{a.s.} (1)$; moreover CMT and (26) $\varsigma_{\min} \left(P_n^2 / n\right) \geq C$, which implies $\left(P_n^2 / n\right)^{-1}$ exists $a.s$. Define $\overline{\phi} = \phi_0 - \left(P_n^2\right)^{-1} Q_n$ and note that (38) can be written as

$$\sum \rho \left(\lambda' \widetilde{m}_t\right) - \sum \rho \left(\lambda' \overline{m}_t\right) = \left(\overline{\phi} - \widetilde{\phi}\right)' P_n^2 \left(\overline{\phi} - \widetilde{\phi}\right) - \left(\overline{\phi} - \phi_0\right)' \overline{R}_n^2 \left(\overline{\phi} - \phi_0\right) + \breve{\xi}_n \tag{39}$$

where

$$\overline{R}_n^2 := \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left[\left(\sum \overline{m}_{tn} \overline{m}_{tn}'\right)^{-1} - \left(\sum m_{tn} m_{tn}'\right)^{-1}\right] \sum Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}'.$$

Note that

$$\left\|\overline{R}_n^2\right\| \leq \left\|\sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}'\right\|^2 \left\|\left(\sum \overline{m}_{tn} \overline{m}_{tn}'\right)^{-1} - \left(\sum m_{tn} m_{tn}'\right)^{-1}\right\|$$

and that $\left\|\left(\sum \overline{m}_{tn}\overline{m}_{tn}'\right)^{-1} - \left(\sum m_{tn}m_{tn}'\right)^{-1}\right\| = o_p(1)$ using (31), (26), the triangle inequality and the same arguments of Berk (1974, p. 493). Since

$$E\left\|\sum \Delta \mathbf{y}_{p,t-1}Y_{t-1,n}'/n^{1/2} - \left[\begin{array}{cc} 0 & \Sigma_p \end{array}\right]'\right\|^2 = O_p(p/n),$$

it follows that $\left\|\sum \Delta \mathbf{y}_{p,t-1}Y_{t-1,n}'\right\|^2 = O_p(p)$ whence by CMT $\left\|\overline{R}_n^2\right\| = O(p) := \delta_n$. Lemma 8 shows that if $p = o\left(n^{1/4}\right)$ $\left\|\sum m_{tn}m_{tn}' - \Sigma_{\omega p}\right\| = o_p(1/p)$ hence $\delta_n = o_p(1)$. Note that for any θ such that $\|\theta\| \leq 1$ $\tau_{pn}^2\theta' P_n^2\theta \geq \varsigma_{\min}\left(P_n^2\right)$,

$$n^{1/2}\left\|\left(\overline{\phi} - \widetilde{\phi}\right)\right\| \leq \varsigma_{\max}\left(\left(P_n^2\right)^{-1}\right)\left\|Q_n/n^{1/2}\right\| = O_p\left(p^{1/2}\right) = o\left(\tau_{pn}\right)$$

and hence $\left(\overline{\phi} - \widetilde{\phi}\right) \in \Phi_{\tau_{pn}}$, and that by triangle inequality $\widetilde{\phi} - \phi_0 \in 2\Phi_{\tau_{pn}}$. Thus since $\sup_{\widetilde{\phi}\in\Phi_{\tau_{pn}}}\left|\left(\overline{\phi} - \phi_0\right)'\overline{R}_n^2\left(\overline{\phi} - \phi_0\right)\right| \leq 4\delta_n\tau_{pn}^2$, and $\sup_{\widetilde{\phi}\in\Phi_{\tau_{pn}}}|\zeta_n| \leq \delta_n\tau_{pn}^2$, we then have that

$$\min \sup_{\widetilde{\phi}\in\Phi_{\tau_{pn}}} \sum \rho\left(\lambda'\widetilde{m}_t\right) - \sum \rho\left(\lambda'\overline{m}_t\right) \geq \left(\varsigma_{\min}\left(P_n^2\right)/2 - 5\delta_n\right)\tau_{pn}^2.$$

Because $\delta_n = o_p(1)$ and $\tau_{pn} > 0$ for each n it follows that $\sum \rho\left(\lambda'\widetilde{m}_t\right)$ attains $a.s.$ its minimum value at some point $\widehat{\phi} \in int\left\{\Phi_{\tau_{pn}}\right\}$ and since $\sum \rho\left(\lambda'\widetilde{m}_t\right)$ is continuous on $\Phi_{\tau_{pn}}$ it follows that $\widehat{\phi}$ satisfies $0 = \sum \partial\rho\left(\lambda'\widetilde{m}_t\right)/\partial\phi$ $a.s.$ ■

Lemma 17 *Assume that M(I),(II), **MU**, **LP**, **P**(I),(II) hold, and let*

$$\widehat{\phi} := \arg\min_{\widetilde{\phi}\in\Phi_{\tau_{pn}}} \sum \rho\left(\lambda'\widetilde{m}_t\right).$$

Then $\left\|\widehat{\phi} - \phi_0\right\| = O_p\left((p/n)^{1/2}\right)$.

Proof. Using Lemma 14 and a Taylor expansion it is possible to show that

$$2\left(\sum \rho\left(\lambda'\widehat{m}_t\right) - \rho(0)\right) = \sum \widehat{m}_{tn}'\left(\sum \widehat{m}_{tn}\widehat{m}_{tn}'\right)^{-1}\sum \widehat{m}_{tn} + o_p(1).$$

Since $\widehat{m}_{tn} = m_{tn} - Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi} - \phi_0\right)$ it follows that

$$\sum \widehat{m}_{tn}'\left(\sum \widehat{m}_{tn}\widehat{m}_{tn}'\right)^{-1}\sum \widehat{m}_{tn} = \sum m_{tn}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum m_{tn} -$$
$$2\sum m_{tn}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi} - \phi_0\right) +$$
$$\left(\widehat{\phi} - \phi_0\right)'\sum \Delta\mathbf{y}_{p,t-1}Y_{t-1,n}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi} - \phi_0\right).$$

Let

$$\left(\widehat{\phi}-\phi_0\right)'\sum \Delta \mathbf{y}_{p,t-1}Y_{t-1,n}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum Y_{t-1,n}\Delta \mathbf{y}_{p,t-1}'\left(\widehat{\phi}-\phi_0\right):=F_n^2,$$

$$\sum \widehat{m}_{tn}'\left(\sum \widehat{m}_{tn}\widehat{m}_{tn}'\right)^{-1}\sum \widehat{m}_{tn}:=\widehat{M}^2,\quad \sum m_{tn}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum m_{tn}:=M_n^2.$$

By repeated use of the triangle inequality we have that

$$F_n^2\leq \widehat{M}_n^2+M_n^2+2F_n'M_n\leq \widehat{M}_n^2+M_n^2+2F_n'\left(\widehat{M}_n+M_n\right)$$

where $F_n'=\left(\widehat{\phi}-\phi_0\right)'\sum \Delta \mathbf{y}_{p,t-1}Y_{t-1,n}'\left(\sum m_{tn}m_{tn}'\right)^{-1/2}$, $M_n'=\sum m_{tn}'\left(\sum m_{tn}m_{tn}'\right)^{-1/2}$ and $\widehat{M}_n$ is defined similarly. Subtracting $2F_n'\left(\widehat{M}_n+M_n\right)$, adding $\left(\widehat{M}_n+M_n\right)^2$ and taking the square roots from both sides yields $\left\|F_n-\left(M_n+\widehat{M}_n\right)\right\|\leq 2^{1/2}\left\|\widehat{M}_n+M_n\right\|$. Note that $M_n^2\leq C\left\|m_{tn}\right\|^2$ and $\widehat{M}_n^2\leq C\left\|\widehat{m}_{tn}\right\|^2$ so that by Lemmae 11, 14 and CMT both M_n^2 and $\widehat{M}_n^2$ are $O_p(p)$. Then again by triangle inequality $\left\|F_n-\left(M_n+\widehat{M}_n\right)\right\|\geq \left\|F_n\right\|-\left\|M_n+\widehat{M}_n\right\|$ hence

$$\left\|F_n\right\|^2\leq C\left\|\widehat{M}_n+M_n\right\|^2=O_p(p).$$

Since $\sum \Delta \mathbf{y}_{p,t-1}Y_{t-1,n}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum Y_{t-1,n}\Delta \mathbf{y}_{p,t-1}'/n\xrightarrow{p}\Sigma_p$ we have that $F_n^2\geq \varsigma_{\min}\left(\Sigma_p\right)n\left\|\widehat{\phi}-\phi_0\right\|^2\geq n\left\|\widehat{\phi}-\phi_0\right\|^2C$ and thus

$$\left\|\widehat{\phi}-\phi_0\right\|^2C\leq F_n^2\leq \widehat{M}_n^2+M_n^2=O_p(p/n).$$

■

Lemma 18 *Let* $y_t=(1-\delta/n)\,y_{t-1}+u_t$ *where* u_t *is as in Lemma 9 and* $y_0=0$. *Then for any* $1\geq\alpha>1/2$ $\max_t|y_t|=o_{a.s.}(n^\alpha)$.

Proof. Let $\beta_{j,t}^\delta=(1-\delta/n)^{t-j}$; by recursive substitution $y_t=\sum_{j=1}^t\beta_{j,t}^\delta u_j$. Using the same notation of Lemma 9, Beveridge-Nelson decomposition and summation by parts give

$$y_t=\beta_{t,t}^\delta\left[\Psi(1)\,S_t+\eta_t\right]-\sum_{j=1}^t S_{j-1}\left(\beta_{j,t}^\delta-\beta_{j-1,t}^\delta\right). \tag{40}$$

By Chow's strong law of large number $S_t = o_{a.s.}(n^\alpha)$ so that by Lemma 9 and triangle inequality

$$\max_t |y_t| \le |\beta^\delta_{t,t}| \max_t |S_t + \zeta_t| + \max_t |S_{t-1}| \sum_{j=1}^{t} |\beta^\delta_{j,t} - \beta^\delta_{j-1,t}| = o_{a.s.}(n^\alpha)$$

since $\sum_{j=1}^{t} |\beta^\delta_{j,t} - \beta^\delta_{j-1,t}| < \infty$. ■

Lemma 19 *Let* $y_t = (1 - \delta/n) y_{t-1} + u_t$ *where* u_t *is as in Lemma 9 and* $y_0 = 0$. *Then*

$$\sum y_{t-1}\varepsilon_{p,t} = \sum y_{t-1}\varepsilon_t + o_p(n).$$

Proof. The proof is similar to the one given in Lemma 3.1 of Chang and Park (2002). For notational convenience let $\varepsilon_{p,t} - \varepsilon_t = \epsilon_{pt}$, and note that $\sum y_{t-1}\varepsilon_{p,t} = \sum y_{t-1}\varepsilon_t + \sum y_{t-1}\epsilon_{pt}$. By (40)

$$\sum y_{t-1}(\varepsilon_{p,t} - \varepsilon_t) = \left[\beta^\delta_{t-1,t-1} \sum \left[\Psi(1) S_{t-1} + \zeta_{t-1}\right] + \sum \sum_{j=2}^{t-1} S_{j-2}\left(\beta^\delta_{j,t-1} - \beta^\delta_{j-1,t-1}\right)\right] \epsilon_{pt}.$$

Consider first $\sum \beta^\delta_{t-1,t-1} S_t \epsilon_{pt}$ and note that $\epsilon_{pt} = \sum_{l=p+1}^{\infty} \gamma_{p,l}\varepsilon_{t-l}$ where $\sum_{l=p+1}^{\infty} \gamma^2_{p,l} \le C \sum_{l=p+1}^{\infty} \alpha_l^2 = o(p^{-2k})$. Let δ_{jk} denote the Kronecker delta; then

$$\begin{aligned}
\sum \beta^\delta_{t-1,t-1} S_{t-1}\epsilon_{pt} &= \sum_{l=p+1}^{\infty} \gamma_{p,l}\beta^\delta_{t-1,t-1} \sum \sum_{i=1}^{t-1} \varepsilon_{t-i}\varepsilon_{t-l} \\
&= \beta^\delta_{t-1,t-1}\left[n\sigma^2 \sum_{l=p+1}^{\infty} |\gamma_{p,l}| + \sum_{l=p+1}^{\infty} \gamma_{p,l} \sum \sum_{i=1}^{t-1} \left(\varepsilon_{t-i}\varepsilon_{t-l} - \sigma^2\delta_{ij}\right)\right] \\
&\le \beta^\delta_{t-1,t-1}\left\{n\sigma^2 \sum_{l=p+1}^{\infty} |\gamma_{p,l}| + \sum_{l=p+1}^{\infty} |\gamma_{p,l}| E^{1/2}\left[\sum \sum_{i=1}^{t-1} \left(\varepsilon_{t-i}\varepsilon_{t-l} - \sigma^2\delta_{il}\right)\right]^2\right\} \\
&\le Co\left(np^{-k}\right).
\end{aligned}$$

Next

$$\begin{aligned}
\sum \beta^\delta_{t-1,t-1}\zeta_{t-1}\epsilon_{pt} &= \sum_{j=0}^{\infty} \alpha_j \sum_{l=p+1}^{\infty} \gamma_{p,l} \sum \beta^\delta_{t-1,t-1}\varepsilon_{t-i-1}\varepsilon_{t-l} \\
&= \beta^\delta_{t-1,t-1}\left[n\sigma^2 \sum_{l=p+1}^{\infty} \alpha_{l-1}\gamma_{p,l} + \sum_{j=0}^{\infty} \alpha_j \sum_{l=p+1}^{\infty} \gamma_{p,l} \sum \left(\varepsilon_{t-i-1}\varepsilon_{t-l} - \sigma^2\delta_{i+1l}\right)\right] \\
&= C\left(o\left(np^{-k}\right) + o_{a.s}\left(n^\eta p^{-k}\right)\right)
\end{aligned}$$

for any $1 \geq \eta > 1/2$ by Chow's strong law. Similarly for the last the term it can be shown that $\sum \sum_{j=2}^{t-1} S_{j-2} \left(\beta_{j,t-1}^{\delta} - \beta_{j-1,t-1}^{\delta}\right) \epsilon_{pt} = o_p\left(np^{-k}\right)$ using $\sum \left|\left(\beta_{j,n}^{\delta} - \beta_{j-1,n}^{\delta}\right)\right| < \infty$. ■

7.4 Proofs of Theorems

Proof of Theorem 1. Lemma 17 shows that $\left\|\widehat{\phi} - \phi_0\right\| = O_p\left((p/n)^{1/2}\right)$, which implies that $\widehat{\phi} - \phi_0 \in int\{\Phi\}$ since $(p/n)^{1/2} = o\left(p^{-1/2}n^{-\beta}\right)$ and hence by Lemma 16 there exists a $\widehat{\phi}$ such that $0 = \partial \sum \rho\left(\lambda' \widehat{m}_t\right)/\partial\phi$ *a.s.* Then by Lemma 15 $\|\widehat{m}_{tn}\| = O_p\left(p^{1/2}\right)$ and hence by Lemma 14 $\widetilde{\lambda} = \arg\max_{\lambda \in \Lambda_0} l(\lambda, \phi)$ exists *a.s.* By Lemma 13 $\max_t \left|\widehat{\lambda}' \widehat{m}_t\right| = o_{a.s.}(1)$ so that for all λ and ϕ in a neighborhood of a neighborhood of $\left(\widehat{\lambda}, \widehat{\phi}\right)$ $\sum \rho\left(\lambda' m_t\right)$ is twice continuously differentiable and $\sum \partial^2 \rho\left(\widehat{\lambda}' \widehat{m}_t\right)/\partial\lambda\partial\lambda' = \sum \rho_2\left(\widehat{\lambda}' \widehat{m}_t\right) \widehat{m}_t \widehat{m}_t'$ and note that $S_n\left[\partial^2 \rho\left(\widehat{\lambda}' \widehat{m}_t\right)/\partial\lambda\partial\lambda'\right] S_n$ is nonsingular *a.s.*. By the implicit function theorem there is a continuously differentiable function $\lambda(\phi)$ such that $0 = \sum \partial\rho\left(\lambda(\phi)' m_t\right)/\partial\phi$ so that by the envelope theorem

$$0 = n^{-1} \sum \partial\rho\left(\widehat{\lambda}' \widehat{m}_t\right)/\partial\phi = n^{-1} \sum \rho_1\left(\widehat{\lambda}_n' \widehat{m}_{tn}\right) \left(\partial\widehat{m}_{tn}/\partial\phi\right)' \widehat{\lambda}_n. \tag{41}$$

Then for $0 \leq \overline{\lambda} \leq \widehat{\lambda}$

$$0 = \sum \partial\rho\left(\widehat{\lambda}' \widehat{m}_t\right)/\partial\lambda = S_n\left[\sum \rho_1(0) \widehat{m}_{tn} + \sum \rho_2\left(\overline{\lambda}' \widehat{m}_t\right) \widehat{m}_{tn} \widehat{m}_{tn}' \widehat{\lambda}_n\right];$$

since by (32) $\max_t \left|\rho_2\left(\widehat{\lambda}' \widehat{m}_t\right) - \rho_2(0)\right| = o_{a.s.}(1)$ it follows by (31), (19) and CMT that $\varsigma_{\min}\left(\sum \rho_2\left(\overline{\lambda}' \widehat{m}_t\right) \widehat{m}_{tn} \widehat{m}_{tn}'\right) > 0$ *a.s.*, and thus $\widehat{\lambda}_n = \left(\sum \widehat{m}_{tn} \widehat{m}_{tn}'\right)^{-1} \sum \widehat{m}_{tn} + o_p(1)$. Using $\widehat{m}_{tn} = m_{tn} - Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}' \left(\widehat{\phi} - \phi_0\right)$ in (41) gives

$$\begin{aligned} 0 &= \sum \left(n^{-1/2} \partial\widehat{m}_{tn}/\partial\phi\right)' \left[\sum \rho_2\left(\overline{\lambda}' \widehat{m}_t\right) \widehat{m}_{tn} \widehat{m}_{tn}'\right]^{-1} \times \\ &\quad \sum \left[m_{tn} - Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}' \left(\widehat{\phi} - \phi_0\right)/n^{1/2}\right] + o_p(1). \end{aligned}$$

Note that $n^{-1/2} \sum \partial\widehat{m}_{tn}/\partial\phi = -\sum Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}'/n^{1/2} \xrightarrow{p} \begin{bmatrix} 0 & -\Sigma_p \end{bmatrix}'$ whereas by (31), (32) and CMT $\left[\sum \rho_2\left(\overline{\lambda}' \widehat{m}_t\right) \widehat{m}_{tn} \widehat{m}_{tn}'\right]^{-1} \xrightarrow{p} -\Sigma_{\omega p}^{-1}$, so that

$$\sum \left(n^{-1/2} \partial\widehat{m}_{tn}/\partial\phi\right)' \left[\sum \rho_2\left(\overline{\lambda}' \widehat{m}_t\right) \widehat{m}_{tn} \widehat{m}_{tn}'\right]^{-1} \sum Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}'/n^{1/2} \xrightarrow{p} \begin{bmatrix} 0 & \Sigma_p \end{bmatrix} \Sigma_{\omega p}^{-1} \begin{bmatrix} 0 & \Sigma_p \end{bmatrix}'$$

and hence $\left(\widehat{\phi}-\phi_0\right)=\Sigma_p^{-1}\sum\Delta\mathbf{y}_{p,t-1}\varepsilon_{tp}$. The latter implies that $\sum\widehat{m}_{tn}=\sum m_{tn}-Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi}-\phi_0\right)=\sum\left[\begin{array}{cc} y_{t-1}\varepsilon_{tp}/n & 0'\end{array}\right]'$ and hence by a further Taylor expansion about $\widehat{\lambda}$

$$2\left(\sum\rho\left(\widehat{\lambda}'\widehat{m}_t\right)-\rho(0)\right) = \sum\widehat{m}_{tn}'\Sigma_{\omega p}^{-1}\sum\widehat{m}_{tn}+o_p(1)=\left(\sum y_{t-1}\varepsilon_{pt}\right)^2/\left(\sigma^2\sum y_{t-1}^2\right)$$
$$\Rightarrow\left(\int_0^1 B(r)\,dB(r)\right)^2/\int_0^1 B^2(r)\,dr. \tag{42}$$

By the duality between GEL and ECR statistics (see (20)) (14) follows. ■

Proof of Theorem 2. Note that

$$\max_t\left|y_{t-1}\left(\varepsilon_{p,t}-\varepsilon_t\right)/n\right|\le\max_t\left|y_{t-1}\varepsilon_t/n\right|\left|\sum_{j=p+1}^{\infty}\psi_j\right|=o_{a.s.}(1)\,o\left(p^{-s}\right)=o_{a.s.}(1)$$

by Lemma 18 and **MU**, so that

$$\max_t\left|y_{t-1}\varepsilon_{p,t}/n\right|\le\max_t\left|y_{t-1}\varepsilon_t/n\right|+\max_t\left|y_{t-1}\left(\varepsilon_{p,t}-\varepsilon_t\right)/n\right|=o_{a.s.}(1). \tag{43}$$

Note that by Phillips (1987*b*) $\sum y_{t-1}^2/n^2\Rightarrow\sigma^2\Psi(1)^2\int_0^1 J_\gamma{}^2(r)\,dr:=\sigma^2\omega_\gamma$ whereas for each $j=1,...,p$ it is possible to show that $\sum\left(y_{t-1}u_{t-j-1}/n\right)^2=O_p(1)$ so that $E\left\|\sum y_{t-1}\Delta\mathbf{y}_{p,t-1}'/n^{3/2}\right\|^2=O(p/n)$. Since $y_t=O_p\left(n^{1/2}\right)$, $\sum\Delta\mathbf{y}_{p,t-1}\Delta\mathbf{y}_{p,t-1}'/n\xrightarrow{p}\Sigma_p$ it then follows that the conclusions of Lemmae 5-7 and hence of Lemma 8 are still valid. Furthermore in view of (43) Lemma 10 is also valid. Thus by the same arguments of Lemma 11 following the proof of Theorem 1 gives

$$2\left(\sum\rho\left(\widehat{\lambda}'m_t\left(\beta_n,\phi_0\right)\right)-\rho(0)\right) = \sum m_{tn}\left(\beta_n,\phi_0\right)'\left(\sum m_{tn}\left(\beta_n,\phi_0\right)m_{tn}\left(\beta_n,\phi_0\right)'\right)^{-1}\times$$
$$\sum m_{tn}\left(\beta_n,\phi_0\right)+o_p(1)$$

Note that $\sum m_{tn}\left(\beta_n,\phi_0\right)m_{tn}\left(\beta_n,\phi_0\right)'\Rightarrow diag\left(\sigma^2\left[\begin{array}{cc}\omega_\gamma & \Sigma_p\end{array}\right]\right)$ and hence ϕ_0 and β are asymptotically independent under the sequence of local alternatives β_n. Thus Lemmae 12-17 can be used exactly as in the proof of Theorem 1 to show that there exists a $\widehat{\phi}\in int\left\{\Phi_{pn}\right\}$ that solves $0=\sum\partial\rho\left(\lambda'm_t\left(\beta_n,\widetilde{\phi}\right)\right)/\partial\phi$. Let $m_{tn}\left(\beta,\widehat{\phi}\right)=Y_{t-1,n}\left(y_t-\beta_n y_{t-1}-\Delta\mathbf{y}_{p,t-1}'\widehat{\phi}\right)$ and note that $\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)=\sum m_{tn}\left(\beta_n,\phi_0\right)-Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi}-\phi_0\right)$. The block diagonality of $\sum m_{tn}\left(\beta_n,\phi_0\right)m_{tn}\left(\beta_n,\phi_0\right)$ implies that $\left(\widehat{\phi}-\phi_0\right)=\Sigma_p^{-1}\Delta \mathbf{y}_{p,t-1}\varepsilon_{tp}$ and hence $\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)=\sum\left[\begin{array}{cc} y_{t-1}\varepsilon_{tp}/n & 0'\end{array}\right]'$ as in Theorem 1. Thus by Taylor expansion

$$\begin{aligned}
& 2\left(\sum\rho\left(\widehat{\lambda}'m_t\left(\beta_n,\phi_0\right)\right)-\rho\left(0\right)\right)=\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)'M_{nn}^{-1}\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)+o_p\left(1\right)\\
= & \sum m_{tn}\left(\beta_n,\widehat{\phi}\right)\left[diag\left(\sigma^2\left[\begin{array}{cc}\omega_\gamma & \Sigma_p\end{array}\right]\right)\right]^{-1}\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)+o_p\left(1\right)\\
= & \left(\sum y_{t-1}\varepsilon_{pt}\right)^2/\left(\sigma^2\sum y_{t-1}^2\right)+o_p\left(1\right).
\end{aligned}$$

By Lemma 19 $\sum y_{t-1}\varepsilon_{pt}/n=\sum y_{t-1}\varepsilon_t/n+o_p\left(1\right)$ so that as in Theorem 1 (15) follows by (20), the results of Phillips (1987*b*) and CMT. ■

Proof of Corollary 3. A straightforward application of CMT shows that $R\left(1,\widehat{\phi},\gamma\right)$ converges weakly to the same distribution as that of the ADF_t statistic. ■

Proof of Theorem 4. The results of Chang and Park (2003) and Park (2002) show that for $S_t^*=\sum_{j=1}^t\varepsilon_j^*$, $\widehat{\Psi}\left(1\right)=1/\left(1-\sum_{j=1}^\infty\widehat{\phi}_k\right)$ and $\widehat{\sigma}^2=\sum\left(\widehat{\varepsilon}_{p(n),t}-E^*\widehat{\varepsilon}_{p(n),t}\right)^2/n$

$$\begin{aligned}
\left(\widehat{\beta}^*-1\right)/\widehat{\sigma}_{\widehat{\beta}^*}^* &= \left(\widehat{\Psi}\left(1\right)\sum S_{t-1}^*\varepsilon_t^*/n\right)/\left(\widehat{\sigma}^2\widehat{\Psi}\left(1\right)\sum S_{t-1}^{*2}/n^2\right)^{1/2}+o_{p^*}\left(1\right) \text{ in probability}\\
&\Rightarrow^* \int_0^1 B\left(r\right)dB\left(r\right)/\left(\int_0^1 B^2\left(r\right)dr\right)^{1/2} \text{ in probability}
\end{aligned}$$

since $\widehat{\sigma}^2\overset{a.s.}{\to}\sigma^2$ by strong law of large numbers. Then by CMT

$$\left(\widehat{\beta}^*-1\right)^2/\widehat{\sigma}_{\widehat{\beta}^*}^{*2}\Rightarrow^*\left(\int_0^1 B\left(r\right)dB\left(r\right)\right)/\left(\int_0^1 B^2\left(r\right)dr\right) \text{ in probability.} \quad (44)$$

Let $ADF_t^{*2}:=\left(\widehat{\beta}^*-1\right)^2/\widehat{\sigma}_{\widehat{\beta}^*}^{*2}$ and for $0<\alpha<1$ $u_{1-\alpha}:=\inf\left\{u:\Pr\left(ADF_t^2\geq u\right)\geq 1-\alpha\right\}$ and $u_{1-\alpha}^*$ denote its bootstrap analogue. From (44) it follows that

$$\Pr{}^*\left(ADF_t^{2*}\geq u_{1-\alpha}\right)\to 1-\alpha \text{ in probability}$$

and thus $u_{1-\alpha}^*\overset{p}{\to}u_{1-\alpha}$, which in turn implies (18). ■

Tables

Table 2. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},0\right) := W_0$, ADF_t^2, $BADF_t^2$ with $AR\ N(0,1)$ errors at 0.10 and 0.05 nominal level $n = 100$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.119	.075	.121	.089	.111	.046	.099	.049
1	-.5	.117	.067	.123	.088	.116	.049	.103	.052
1	-.2	.115	.061	.119	.065	.109	.045	.101	.051
1	0	.112	.067	.115	.072	.120	.070	.105	.056
1	.2	.116	.059	.120	.068	.109	.049	.101	.052
1	.5	.121	.078	.125	.080	.120	.073	.105	.054
1	.8	.115	.069	.117	.075	.110	.063	.101	.051
.95	-.8	.427	.288	.425	.296	.408	.253	.332	.212
.95	-.5	.449	.315	.464	.331	.414	.241	.301	.218
.95	-.2	.436	.305	.455	.318	.383	.226	.273	.210
.95	0	.409	.256	.427	.267	.399	.204	.226	.168
.95	.2	.388	.312	.401	.298	.366	.236	.264	.187
.95	.5	.377	.216	.379	.237	.301	.217	.221	.173
.95	.8	.312	.217	.331	.213	.305	.179	.214	.165
.90	-.8	.723	.535	.733	.531	.625	.527	.576	.487
.90	-.5	.750	.615	.759	.626	.750	.556	.612	.499
.90	-.2	.738	.629	.757	.641	.755	.543	.689	.497
.90	0	.697	.545	.678	.534	.705	.598	.677	.534
.90	.2	.709	.515	.728	.533	.639	.522	.556	.451
.90	.5	.637	.423	.682	.483	.475	.396	.327	.298
.90	.8	.505	.365	.537	.373	.446	.336	.303	.263

Table 3. Finite sample size and power of $W\left(1, \widehat{\phi}, -1\right) := W_{-1}$, $W\left(1, \widehat{\phi}, 0\right) := W_0$, ADF_t^2, $BADF_t^2$ with $MA\ N(0,1)$ errors at 0.10 and 0.05 nominal level $n = 100$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.124	.075	.131	.089	.125	.069	.108	.059
1	-.5	.119	.059	.125	.076	.110	.060	.105	.052
1	-.2	.106	.057	.099	.062	.101	.056	.099	.050
1	0	.112	.069	.110	.073	.113	.064	.104	.051
1	.2	.107	.064	.109	.068	.113	.064	.104	.056
1	.5	.104	.054	.113	.067	.109	.056	.102	.053
1	.8	.107	.063	.113	.070	.111	.061	.102	.053
.95	-.8	.554	.356	.592	.373	.524	.343	.475	.303
.95	-.5	.421	.289	.438	.296	.448	.225	.406	.211
.95	-.2	.489	.354	.466	.378	.517	.346	.423	278
.95	0	.357	.234	.399	.217	.368	.199	.299	.156
.95	.2	.443	234	.472	.250	.462	.267	.387	.202
.95	.5	.432	.201	.443	.228	.438	.265	.346	.123
.95	.8	.321	.150	.347	.174	.321	.172	.267	.120
.90	-.8	.769	.650	.759	.670	.880	.687	.744	604
.90	-.5	.798	.614	.733	.594	.887	.676	.703	.597
.90	-.2	.759	.540	.760	.583	.853	.688	.698	.578
.90	0	.709	.453	.763	.498	.856	.597	.674	.502
.90	.2	.699	.379	.715	.412	.744	.548	.645	.432
.90	.5	.754	.384	.797	.404	.768	.559	.603	.443
.90	.8	.590	.364	.601	.382	.564	.348	.503	.302

Table 4. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},-1\right) := W_0$, ADF_t^2, $BADF_t^2$ with AR χ_4^2-4 errors at 0.10 and 0.05 nominal level $n=100$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.126	.107	.133	.103	.110	.060	.106	.053
1	-.5	.115	.064	.110	.055	.121	.071	.104	.054
1	-.2	.115	.070	.116	.067	.105	.051	.101	.052
1	0	.107	.65	.108	.060	.108	.052	.102	.051
1	.2	.110	.069	.116	.078	.102	.050	.100	.048
1	.5	.117	.073	.121	.080	.118	.063	.105	.054
1	.8	.115	.072	.130	.078	.118	.068	.104	.055
.95	-.8	.411	.305	.409	.314	.357	.245	.296	.198
.95	-.5	.420	.329	.426	.334	.402	.215	.333	.206
.95	-.2	.446	.347	.465	.350	.396	.240	.307	.205
.95	0	.443	.327	.455	.340	.377	.312	.313	.197
.95	.2	.446	.343	.463	.357	.397	.233	.325	.169
.95	.5	.409	.317	.417	.325	.399	.246	.329	.157
.95	.8	.398	.248	.385	.278	.336	.182	.245	.136
.90	-.8	.699	.508	.714	.514	.545	.483	.435	.377
.90	-.5	.734	.537	.750	.571	.572	.519	.517	.443
.90	-.2	.714	.603	.748	.630	.697	.538	.521	.437
.90	0	.732	.599	.740	.605	.655	.560	.553	.430
.90	.2	.713	.581	.731	.587	.672	.533	.604	.424
.90	.5	.655	.518	.673	.510	.608	.463	.523	.421
.90	.8	.514	.405	.523	.413	.528	.341	.424	.321

Table 5. Finite sample size and power of $W\left(1, \widehat{\phi}, -1\right) := W_{-1}$, $W\left(1, \widehat{\phi}, -1\right) := W_0$, ADF_t^2, $BADF_t^2$ with MA $\chi_4^2 - 4$ errors at 0.10 and 0.05 nominal level $n = 100$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.118	.070	.123	.075	.134	.068	.104	.053
1	-.5	.119	.072	.128	.081	.105	.057	.101	.052
1	-.2	.120	.075	.132	.083	.114	.064	.105	.058
1	0	.130	.085	.147	.105	.115	.061	.103	.055
1	.2	.122	.084	.126	.091	.114	.062	.102	.054
1	.5	.117	.076	.122	.08	.116	.061	.108	.054
1	.8	.120	.088	.137	.094	.120	.069	.105	.051
.95	-.8	.599	.401	.637	.425	.620	.427	.543	.354
.95	-.5	.405	.312	.446	.305	.454	.285	.324	177
.95	-.2	.364	.215	.397	.236	.404	.202	.304	.156
.95	0	.437	.230	.468	.246	.418	.257	.323	185
.95	.2	.421	.224	.439	.238	.410	.259	.302	.201
.95	.5	.340	.184	.352	.197	.366	.204	.310	.156
.95	.8	.280	.175	.312	.184	.287	.168	.205	.103
.90	-.8	.785	.626	.744	.682	.863	.649	.704	.535
.90	-.5	.767	.649	.737	.631	885	.704	.712	.597
.90	-.2	.759	.455	.750	.468	.720	.499	.648	.407
.90	0	.742	.537	.788	.525	.737	.555	.659	.438
.90	.2	.708	.503	.738	.597	.719	.523	.620	.465
.90	.5	.712	.434	.792	.412	.660	.447	.539	.367
.90	.8	.533	.403	.572	.463	.554	.403	.501	.325

Table 6. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},-1\right) := W_0$, ADF_t^2, $BADF_t^2$ with AR t_5 errors at 0.10 and 0.05 nominal level $n = 100$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.124	.092	.136	.115	.109	.052	.105	.050
1	-.5	.118	.083	.134	.094	.110	.053	.104	.052
1	-.2	.109	.059	.112	.063	.114	.062	.104	.053
1	0	.110	.060	.107	.065	.109	.058	.103	.049
1	.2	.108	.064	.111	.073	.116	.067	.105	.052
1	.5	.115	.071	.120	.080	.104	.077	.103	.055
1	.8	.117	.066	.125	.073	.121	.064	.106	.054
.95	-.8	.367	.288	.380	.286	.364	.218	.276	.177
.95	-.5	.487	.379	.507	.393	.415	.262	.345	.198
.95	-.2	.469	.355	.482	.362	.389	.230	.317	.184
.95	0	.436	.329	.445	.325	.404	.236	.327	.198
.95	.2	.418	.307	.425	.325	.421	.244	.344	.205
.95	.5	.367	.257	.387	.267	.361	.201	.298	.186
.95	.8	.344	.298	.345	.311	.244	.177	.198	.169
.90	-.8	.689	.527	.706	.540	.589	.503	.439	396
.90	-.5	.734	.633	.776	.652	.750	.602	.654	.552
.90	-.2	.736	.607	.752	.634	.741	.561	.635	.504
.90	0	.721	.599	.749	.605	.703	.543	.619	.454
.90	.2	.709	.558	.726	.579	.658	.556	.549	.446
.90	.5	.687	.456	.684	.487	.521	.447	.432	.388
.90	.8	.487	.356	.501	.367	463	.323	.396	.276

Table 7. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},-1\right) := W_0$, ADF_t^2, $BADF_t^2$ with MA t_5 errors at 0.10 and 0.05 nominal level $n = 100$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.123	.90	.143	.105	.118	.064	.103	.055
1	-.5	.119	.076	.128	.088	.111	.059	.105	.054
1	-.2	.109	.057	.119	.069	.110	.051	.103	.052
1	0	.107	.058	.113	.072	.106	.053	.101	.053
1	.2	.109	.061	.112	.069	.106	.057	.104	.053
1	.5	.110	.061	.121	.076	.112	.058	.105	.052
1	.8	.114	.078	.121	.082	.119	.067	.108	.054
.95	-.8	.498	.313	.522	.334	.573	.353	.476	.288
.95	-.5	.437	.250	.455	.273	.446	.255	.375	.202
.95	-.2	.422	.217	.437	.231	.424	.222	.337	.195
.95	0	.343	.240	.389	.243	.373	.212	.316	.156
.95	.2	.306	.280	.343	.272	.412	.273	.343	.211
.95	.5	.366	.200	.392	.212	.372	.232	.301	.1597
.95	.8	.359	.159	.383	.162	.291	.153	.204	.104
.90	-.8	.813	.680	.842	.694	.896	.679	.786	.589
.90	-.5	.810	.656	.837	.638	.883	.653	.750	.634
.90	-.2	.756	.489	.779	.458	.722	.501	.660	.389
.90	0	.629	.413	.663	.432	.721	.425	.624	.423
.90	.2	.720	.429	.723	.432	.743	.516	.612	.432
.90	.5	.733	.343	.763	.365	.721	.504	.605	.412
.90	.8	.649	.344	.692	.371	.591	.352	.423	.299

Table 8. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right):=W_{-1}$, $W\left(1,\widehat{\phi},0\right):=W_0$, ADF_t^2 and $BADF_t^2$ with $AR\ N(0,1)$ errors at 0.10 and 0.05 nominal level $n=200$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	−.8	.109	.055	.116	.053	.113	.054	.102	.052
1	−.5	.110	.060	.108	.057	.109	.052	.101	.053
1	−.2	.105	.054	.108	.047	.110	.048	.103	.051
1	0	.109	.055	.105	.052	.107	.053	.102	.050
1	.2	.115	.060	.110	.058	.108	.064	.105	.055
1	.5	.117	.061	.112	.059	.109	.056	.104	.052
1	.8	.110	.059	.106	.057	.107	.055	.103	.049
.95	−.8	.799	.630	.805	.643	.811	.617	.784	.600
.95	−.5	.839	.643	.832	.666	.836	.652	.802	.630
.95	−.2	.786	.604	.800	.623	.791	.601	.793	.589
.95	0	.799	.588	.804	.603	.795	.591	.775	.594
.95	.2	.789	.625	.806	.618	.791	.613	.743	.589
.95	.5	.769	.539	.799	.569	.752	.524	.743	.504
.95	.8	.672	.489	.687	.508	.689	.492	.655	.437
.90	−.8	.976	.920	.988	.948	.996	.972	.900	.845
.90	−.5	.971	.912	.988	.970	.994	.964	.884	.856
.90	−.2	.974	.915	.990	.956	.987	.943	.820	.798
.90	0	.954	.937	.984	.955	.985	.947	.832	.812
.90	.2	.943	.929	.991	.964	.989	.956	.856	.829
.90	.5	.952	.905	.989	.934	.981	.916	.877	.841
.90	.8	.989	.959	.990	.968	.994	.964	.899	.869

Table 9. Finite sample size and power of $W\left(1,\widehat{\phi},0\right) := W_0$, $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, ADF_t^2 and $BADF_t^2$ with $MA\ N(0,1)$ errors at 0.10 and 0.05 nominal level $n = 200$

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	−.8	.129	.075	.127	.080	.123	.064	.106	.055
1	−.5	.115	.070	.120	.069	.115	.054	.103	.052
1	−.2	.113	.066	.118	.069	.120	.063	.105	.054
1	0	.125	.075	.120	.073	.117	.062	.108	.053
1	.2	.129	.068	.131	.075	.125	.064	.107	.058
1	.5	.123	.070	.128	.073	.122	.066	.109	.057
1	.8	.137	.076	.132	.082	.124	.064	.105	.059
.95	−.8	.599	.350	.605	.373	.524	.343	.480	.289
.95	−.5	.400	.205	.408	.232	.368	.207	.314	.197
.95	−.2	.320	.210	.357	.208	.378	.228	.307	.199
.95	0	.343	.195	.399	.217	.368	.199	.319	178
.95	.2	.375	.165	.389	.144	.397	.174	.336	.168
.95	.5	.329	.205	.367	.216	.339	.198	.287	.187
.95	.8	.300	.195	.274	.204	.311	.178	.254	.155
.90	−.8	.860	.676	.889	.690	.880	.687	.750	.643
.90	−.5	.643	.496	.627	.485	.688	.503	.605	.487
.90	−.2	.599	.406	.615	.433	.606	.416	.543	.398
.90	0	.587	.395	.594	.404	.590	.388	.504	.376
.90	.2	.465	.370	.493	.399	.568	.376	.523	.342
.90	.5	.553	.390	.580	.396	.568	.397	.536	.354
.90	.8	.439	.299	.465	.304	.457	.298	.397	.275

Table 10. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},0\right) := W_0$, ADF_t^2, $BADF_t^2$ with AR χ_4^2-4 errors at 0.10 and 0.05 nominal level

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	−.8	.109	.055	.107	.057	.102	.053	.104	.051
1	−.5	.108	.057	.107	.061	.106	.053	.102	.052
1	−.2	.110	.055	.105	.060	.104	.056	.102	.053
1	0	.109	.060	.106	.057	.102	.052	.104	.053
1	.2	.113	.056	.110	.058	.104	.055	.102	.052
1	.5	.105	.057	.107	.058	.104	.052	.102	.053
1	.8	.108	.059	.112	.062	.107	.059	.103	.054
.95	.-8	.806	.670	.832	.650	.812	.628	.786	.604
.95	-.5	.796	.606	.822	.612	.792	.595	.775	.557
.95	-.2	.805	.603	.805	.633	.797	.616	.722	.574
.95	0	.757	.598	.799	.610	.778	.582	.732	.532
.95	.2	.756	.598	.800	.603	.781	.581	.700	.554
.95	.5	.743	.567	.764	.576	.750	.549	.698	.504
.95	.8	.676	.500	.696	.496	.683	.479	.632	.443
.90	-.8	.988	.923	.976	.942	.994	.957	.930	.902
.90	-.5	.987	.920	.965	.932	.995	.955	.912	.943
.90	-.2	.960	.932	.945	.940	.994	.959	.921	.903
.90	0	.950	.926	.958	.930	.984	.938	.913	.890
.90	.2	.967	.932	.945	.921	.988	.948	.904	860
.90	.5	.955	.930	.968	.925	.981	.928	.914	.856
.90	.8	.905	.820	.914	.815	.921	.804	.855	.765

Table 11. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},0\right) := W_0$, ADF_t^2 and ADF_t^2 with MA $\chi_4^2 - 4$ errors at 0.10 and 0.05 nominal level

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.119	.065	.123	.070	.124	.068	.104	.053
1	-.5	.115	.075	.124	.086	.120	.065	.102	.054
1	-.2	.109	.069	.118	.077	.122	.067	.105	.053
1	0	.107	.061	.115	.068	.111	.065	.107	.054
1	.2	.110	.069	.128	.075	.125	.070	.106	.056
1	.5	.110	.073	.114	.077	.109	.060	.102	.052
1	.8	.105	.070	.131	.089	.117	.069	.105	.051
.95	.-8	.567	.417	.598	.487	.589	.448	.407	.335
.95	-.5	.386	.240	.393	.285	.377	.224	.305	.198
.95	-.2	.377	.208	.400	.215	.397	.205	.321	.205
.95	0	.363	.202	.385	.199	.355	.195	.238	.176
.95	.2	.380	.185	.379	.194	.343	.178	.277	.143
.95	.5	.357	.213	.363	.201	.358	.197	.310	.155
.95	.8	.377	.244	.388	.213	.342	.187	.300	.156
.90	-.8	.836	.644	.843	.675	.863	.669	.777	.614
.90	-.5	.699	.433	.712	.423	.703	.446	.633	.367
.90	-.2	.598	.359	.604	.378	.587	.345	.503	.302
.90	0	.539	.370	.543	.367	.567	.344	.479	.314
.90	.2	.566	.375	.586	.377	.575	.369	.487	.307
.90	.5	.579	.388	.592	.392	.580	.362	.508	.309
.90	.8	.503	.332	.525	.354	.530	.336	.479	.303

Table 12. Finite sample size and power of $W\left(1,\widehat{\phi},-1\right) := W_{-1}$, $W\left(1,\widehat{\phi},0\right) := W_0$, ADF_t^2 and $BADF_t^2$ with AR t_5 errors at 0.10 and 0.05 nominal level

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.120	.087	.143	.105	.118	.064	.106	.053
1	-.5	.110	.057	.109	.060	.108	.058	.104	.052
1	-.2	.103	.058	.104	.059	.103	.056	.101	.051
1	0	.102	.054	.109	.056	.105	.054	.100	.053
1	.2	.105	.057	.111	.055	.105	.053	.104	.049
1	.5	.110	.057	.108	.059	.104	.052	.103	.048
1	.8	.105	.060	.113	.065	.111	.057	.103	.053
.95	-.8	.960	.906	.980	.945	.831	.635	.744	.600
.95	-.5	.840	.700	.850	.649	.809	.625	.725	.567
.95	-.2	.832	.612	.847	.605	.804	.591	.704	.557
.95	0	.813	.598	.821	.610	.802	.602	.735	.523
.95	.2	.800	.605	.831	.623	.800	.597	.722	.513
.95	.5	.742	.543	.745	.588	.739	.532	.715	.521
.95	.8	.677	.507	.708	.513	.686	.485	.637	.423
.90	-.8	.940	.933	.980	.945	.991	.963	.840	.743
.90	-.5	.956	.934	.990	.966	.992	.958	.853	.714
.90	-.2	.944	.905	.983	.976	.949	.987	.814	.789
.90	0	.955	.930	.967	.960	.983	.959	.855	.810
.90	.2	.959	.905	.967	.932	.985	.947	.843	.809
.90	.5	.978	.887	.974	.921	.988	.939	.832	.738
.90	.8	.932	.745	.915	.788	.922	.779	.834	.754

Table 13. Finite sample size and power of $W\left(1,\widetilde{\phi},-1\right) := W_{-1}$, $W\left(1,\widetilde{\phi},0\right) := W_0$, ADF_t^2 and $BADF_t^2$ with MA t_5 errors at 0.10 and 0.05 nominal level

β	θ	W_{-1}		W_0		ADF_t^2		$BADF_t^2$	
1	-.8	.120	.070	.123	.076	.118	.064	.103	.055
1	-.5	.115	.073	.118	.084	.113	.059	.102	.053
1	-.2	.124	.068	.125	.074	.114	.060	.105	.054
1	0	.129	.074	.136	.079	.113	.065	.103	.059
1	.2	.115	.069	.122	.073	.112	.065	.106	.057
1	.5	.113	.075	.114	.087	.119	.073	.106	.054
1	.8	.124	.065	.129	.78	.123	.076	.104	.053
.95	-.8	.565	.376	.597	.393	.593	.383	.504	245
.95	-.5	.403	.289	.402	.294	.389	.275	.312	.188
.95	-.2	.375	.205	.383	.212	.343	.194	.255	.124
.95	0	.335	.197	.368	.207	.339	.199	.243	.139
.95	.2	.322	.212	.343	.235	.364	.207	.215	.132
.95	.5	.387	.154	.394	.163	.379	.152	.242	.105
.95	.8	.367	.177	.387	.173	.389	.162	.269	.120
.90	-.8	.704	.567	.742	.594	.896	.679	.786	.553
.90	-.5	.589	.456	.605	.472	.630	.458	.589	.368
.90	-.2	.556	.388	.573	.405	.529	.396	.503	.337
.90	0	.576	.387	.582	.397	.599	.384	.534	.318
.90	.2	.534	.354	.567	.374	.572	.371	.516	.312
.90	.5	.512	.314	.554	.342	.574	.355	.507	.298
.90	.8	.521	.299	.543	.304	.549	.315	.492	.275